DEPARTMENT OF ELECTRONIC & ELECTRICAL ENGINEERING

University of Strathclyde Engineering

Gaussian Process Models for SCADA Data Based Wind Turbine Performance/Condition Monitoring

Ravi Kumar Pandit

A thesis submitted for the degree of Doctor of Philosophy
Department of Electronic and Electrical Engineering
University of Strathclyde
2018

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by the University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date:

Abstract

Wind energy has seen remarkable growth in the past decade, and installed wind turbine capacity is increasing significantly every year around the globe. The presence of an excellent offshore wind resource and the need to reduce carbon emissions from electricity generation are driving policy to increase offshore wind generation capacity in UK waters. Logistic and transport issues make offshore maintenance costlier than onshore and availability correspondingly lower, and as a result, there is a growing interest in wind turbine condition monitoring allowing condition based, rather than corrective or scheduled, maintenance.

Offshore wind turbine manufacturers are constantly increasing the rated size the turbines, and also their hub height in order to access higher wind speeds with lower turbulence. However, such scaling up leads to significant increments in terms of materials for both tower structure and foundations, and also costs required for transportation, installation, and maintenance. Wind turbines are costly affairs that comprise several complex systems connected altogether (e.g., hub, drive shaft, gearbox, generator, yaw system, electric drive and so on). The unexpected failure of these components can cause significant machine unavailability and/or damage to other components. This ultimately increases the operation and maintenance (O&M) cost and subsequently cost of energy (COE). Therefore, identifying faults at an early stage before catastrophic damage occurs is the primary objective associated with wind turbine condition monitoring.

Existing wind turbine condition monitoring strategies, for example, vibration signal analysis and oil debris detection, require costly sensors. The additional costs can be significant depending upon the number of wind turbines typically deployed in offshore wind farms and also, costly expertise is generally required to interpret the results. By contrast, Supervisory Control and Data Acquisition

(SCADA) data analysis based condition monitoring could underpin condition based maintenance with little or no additional cost to the wind farm operator.

A Gaussian process (GP) is a stochastic, nonlinear and nonparametric model whose distribution function is the joint distribution of a collection of random variables; it is widely suitable for classification and regression problems. GP is a machine learning algorithm that uses a measure of similarity between subsequent data points (via covariance functions) to fit and or estimate the future value from a training dataset. GP models have been applied to numerous multivariate and multi-task problems including spatial and spatiotemporal contexts. Furthermore, GP models have been applied to electricity price and residential probabilistic load forecasting, solar power forecasting. However, the application of GPs to wind turbine condition monitoring has to date been limited and not much explored.

This thesis focuses on GP based wind turbine condition monitoring that utilises data from SCADA systems exclusively. The selection of the covariance function greatly influences GP model accuracy. A comparative analysis of different covariance functions for GP models is presented with an in-depth analysis of popularly used stationary covariance functions. Based on this analysis, a suitable covariance function is selected for constructing a GP model-based fault detection algorithm for wind turbine condition monitoring.

By comparing incoming operational SCADA data, effective component condition indicators can be derived where the reference model is based on SCADA data from a healthy turbine constructed and compared against incoming data from a faulty turbine. In this thesis, a GP algorithm is constructed with suitable covariance function to detect incipient turbine operational faults or failures before they result in catastrophic damage so that preventative maintenance can be scheduled in a timely manner. In order to judge GP model effectiveness, two other methods, based on binning, have been tested and compared with the GP based algorithm. This thesis also considers a range of critical turbine parameters and their impact on the GP fault detection algorithm.

Power is well known to be influenced by air density, and this is reflected in the IEC Standard air density correction procedure. Hence, the proper selection of an air density correction approach can improve the power curve model. This thesis addresses this, explores the different types of air density correction approach, and suggests the best way to incorporate these in the GP models to improve accuracy and reduce uncertainty.

Finally, a SCADA data based fault detection algorithm is constructed to detect failures caused by the yaw misalignment. Two fault detection algorithms based on IEC binning methods (widely used within the wind industry) are developed to assess the performance of the GP based fault detection algorithm in terms of their capability to detect in advance (and by how much) signs of failure, and also their false positive rate by making use of extensive SCADA data and turbine fault and repair logs. GP models are robust in identifying early anomalies/failures that cause the wind turbine to underperform. This early detection is helpful in preventing machines to reach the catastrophic stage and allow enough time to undertake scheduled maintenance, which ultimately reduces the O&M, cost and maximises the power performance of wind turbines. Overall, results demonstrate the effectiveness of the GP algorithm in improving the performance of wind turbines through condition monitoring.

Acknowledgements

First and foremost, I would like to express my greatest gratitude to Professor **David Infield** for his thoughtful academic guidance and positive encouragement throughout the duration of my research. He was patient and gently suggested a series of ideas which worked and led to the success of this PhD research. He has been supportive since the days I began working on this PhD project. I still remember how he helped me in my PhD research and used to motivate me if research paper got rejected and used to say something like 'Don't worry, Ravi, you should take this rejection as positive feedback and improve your research qualities and submit again' to encourage me not to get mentally disturbed by these failures. Though he left the University of Strathclyde at the end of December 2017, still, he is helping me in my PhD thesis writing and giving his valuable advice and moral support. For this again, I express my greatest gratitude.

Dr David McMillan and **Dr Alasdair McDonald** are my current supervisors who continue to provide me their academic guidance and positive encouragement. It was enjoyable and overwhelming working experience with them, and for that, I would like to thank them from my bottom of my heart.

My sincere thanks also goes to **Dr Peter Clive** from the **Wood Group** (previously known as Sgurr Energy) and **Dr Emilio Gomez-Lazaro** from **University of Castilla-La Mancha (UCLM)** who provided me an opportunity to join their team as Visiting Researcher, and who gave access to their research facilities. Without their precious support, it would not be possible to conduct this PhD research.

I also want to take this opportunity to give my special thanks to the researchers and academics in the department for their valuable and helpful suggestions to my work. Particular thanks are due to Edward Hart who helped to clear my doubt.

I would also like to thank by gratefully acknowledging the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 642108 which funded this work.

I am grateful to Drew Smith (Wind Energy Systems CDT administrative staff) for making my PhD experience immensely enjoyable and smooth at Electronic and Electrical Department, University of Strathclyde.

Finally, I want to dedicate this work to my wife who supported and motivated me throughout this challenging period of my life. Special thanks also to my dad, mom, ma (mother in law) and all my family's members and friends for their unconditional love and support.

Above all, I owe it all to Sri Ram (Almighty God) for granting me the wisdom, health, and strength to undertake this research task and enabling me to its completion

Table of Contents

List of Figures

List of Tables

List of abbreviations

ANNs	Artificial Neural Networks
AE	Acoustic Emission
BNs	Bayesian Networks
BOP	Balance Of Plant
CAPEX	Capital Expenditure
CIs	Confidence Intervals
CM	Condition Monitoring
CMS	Condition Monitoring Systems
CART	Classification and Regression Trees
DBN	Dynamic Bayesian Network
DFIG	Double-fed Induction Generator
GP	Gaussian Process
GPR	Gaussian Process regression
GMM	Gaussian Mixture Model
kNN	k-Nearest Neighbour
NSET	Nonlinear State Estimation Technique
OEMs	Original Equipment Manufacturers
OPEX	Operational Expenditure
O&M	Operation & Maintenance
PDF	Probability Distribution Function
RMSE	Root Mean Square Error
RBF	Radial basis function
MAE	Mean Absolute Error
NSET	Nonlinear State Estimation Technique
NMAPE	Normalized mean absolute percentage error
RVs	Random Variables
SCADA	Supervisory Control and Data Acquisition
sMAPE	symmetric mean absolute percentage error
SOFM	Self-organising feature map
SNR	Signal-to-noise ratio
RF	Random Forest
RQ	Rational Quadratic
SVM	Support Vector Machine
SEMS	self-evolving maintenance scheduler
KKT	Karush-Kuhn-Tucker
WTs	Wind Turbines
WTGS	wind turbine generator system

List of variables

A	Swept area (m^2)
K	Covariance matrix
C_v	Modified Bessel function
C_p	Power coefficient of the wind turbine
ECov	Exponential covariance function
SECov	Square Exponential covariance function
μ	Mean function
$\sum$	Covariance function
Γ	Gamma function
$\overline{f*}$	Mean function of the Gaussian Process
l	Characteristic length scale
v	non-negative parameter
n	number of input parameters
$Var[f*]$	The variance of the Gaussian Process
σ_f^2	Signal variance
σ_n^2	Noise variance
k_{RQ}	The rational quadratic covariance function
ρ	Air density (kg/m^3)
v	Hub height wind speed (m/sec)
λ	Tip speed ratio
β	Pitch angle
H_o	null hypothesis
R^2	Coefficient of determination

Chapter 1

Introduction and research motivation; aim and objectives

Due to the presence of an abundant wind resource in many countries, wind energy has become one of the most promising renewable energy resources for power generation. Wind turbine technology has steadily improved, but despite this, there is a pressing need to make the wind industry more profitable; a fundamental way to do this is to reduce operation and maintenance (O&M) costs. According to [1], O&M makeup 20% -25% of the total lifetime costs of an offshore wind farm and this percentage can reach 35% for a wind turbine at the end of life. This highlights the importance of research into O&M in order to reduce these significant costs. Unexpected failures of turbine components make O&M expensive and reduce reliability and energy yield.

In this chapter, Section 1.1 outlines the research challenge that this thesis attempts to address. Section 1.2 details the aim and objectives of the work. In Section 1.3, the novel contributions of this PhD thesis are summarised. Section 1.4 provides an overview of the thesis, briefly describing the content of each chapter and the general thesis layout. Section 1.5 lists the research publications that have resulted from the research undertaken for the thesis.

1.1 Problem description/ Motivation

With the rapid increase of wind farm installation, it is becoming more essential to analyse the performance of operational wind turbines in order to maximise their efficiency in the long run and produce a maximum power output. The increasing trend for offshore deployment plus poor offshore availability and accessibility suggest the need for an effective and reliable condition based maintenance scheme for wind turbines, rather than the traditional condition monitoring techniques, deployed on wind turbines (onshore and offshore) across Europe and America which can support a responsive repair and scheduled maintenance strategy. A recent trend indicates an increasing failure rate as wind turbine capacity rises. This is due to the complexity of the wind turbine system. The use of expensive sensors increase the overall cost of turbines.

The P-F curve (shown in Figure 1.1) illustrates the behaviour of the component before functional failure has occurred [2]. It is worth noting that the P-F curve may take various shapes, linear or exponential, but is generally represented as an exponential as shown in Figure 1.1. The horizontal axis of the P-F curve defines time-in-service for the machine, and the vertical axis signifies the performance measure.

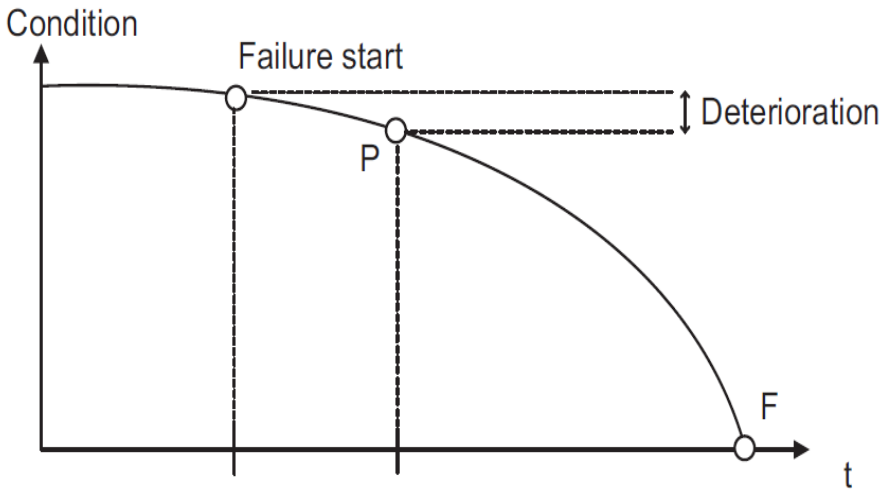


Figure 1.1: Evolution of component condition with deterioration [2]

In Figure 1.1, P stands for the Point of Defect where potential failure of the asset can be identified and F stands for the point of Functional Failure or Loss of Function and the 'window' of time between these points is crucial for effective deployment of condition-based maintenance since it allows for improved maintenance scheduling. From an economic point of view, early detection of a potential anomaly is necessary which prevent catastrophic failure and resulting downtime. For example, a lack of attention to a worn bearing can result in a catastrophic failure of an entire gearbox involving considerable financial loss and significant downtime. This leads to the potential benefits of advanced up-to-dated condition monitoring approaches for minimising breakdown while saving costs by avoiding periodic assessment and associated downtime in context to wind turbine O&M.

Approaches that can identify faults at an early stage correctly without false alarms gives sufficient time for the turbine operator to make decisions regarding maintenance scheduling before the incipient anomalies reach a catastrophic stage, would, therefore, be beneficial for reducing turbine downtime and improving the availability.

This thesis presents a machine learning approach for wind turbine condition monitoring based on an operational power curve where 10-minute SCADA data is used. The specific objectives associated with this thesis are outlined in the next section.

1.2 Aims and Objectives

The main aim of the work set out in this thesis was to investigate and explore the application of Gaussian Process (GP) models for wind turbine condition monitoring based on SCADA data analysis. While doing so, different methodologies were taken into account in order to improve the performance of GP models and come out with a GP model that requires low processing power and low computational cost. Since most damages due to wind turbine unexpected

failures lead to catastrophic failures causes' significant downtime and high maintenance cost. Particular emphasis has been to GP models in this study due to their potential to detect such failures quickly with no or few false alarms. The overall objectives associated with this PhD work are:

- To compare the available range of statistical and machine learning based analysis methods and assess which of these are most effective at extracting wind turbine component condition from wind farm SCADA data. Advantages and disadvantages of the different approaches will be identified.
- The assess the performance of the leading SCADA data analysis algorithms in terms of their capability to detect in advance (and by how much) signs of failure, and also their false positive rate by making use of extensive SCADA data and turbine fault and repair logs.
- To improve upon existing SCADA data analysis algorithms and develop new approaches where appropriate, and to test and refine these in relation to extensive operational data.

In short, this PhD project is focused on SCADA based wind turbine condition monitoring using GP where the full potential of GP models is explored. Particular emphasis is given to improving GP model accuracy by incorporating key variables relevant to wind turbine operation.

1.3 Contribution to knowledge

This thesis has attempted to introduce practical, quick and convenient GP models to assess the performance and or condition of an operating wind farm via the use of SCADA and modelled wind data; The developed GP models work either as an independent performance assessment tool or as a complementary tool for improving the condition monitoring system of wind turbines. The research problem outline for this thesis is as follow.

Can GP models improve SCADA based performance/condition monitoring?

To address this research problem, the following original contributions to knowledge have been made by the author; further details can be found in Chapters 3 to 8 and are summarised as follows in chronological orders:

- Covariance function is used to describe the similarity between two points and its impact the GP models accuracy. Therefore, it is necessary to select an appropriate covariance function for condition monitoring purposes. GP power curve model constructed based on different types of popularly used covariance function and then the comparative analysis is being carried out in order to find out the most suitable covariance function for GP models.
- Two advanced nonparametric approaches, namely, Random Forest (RF), and Support Vector Machine (SVM) are undertaken in order to compare GP model performance (concerning uncertainty and accuracy) where power curve used as a key indicator to assess the individual methods modelling accuracy. Furthermore, results obtained from these methods have been compared using suitable performance error metrics to find out the best method for modelling and robust fault detection capabilities based power curve.
- SCADA based three operational curves constructed using GP model for effective condition monitoring. These GP operational curves (power curve, rotor curve and blade pitch angle curve) are then compared with IEC binning method (widely used by wind industry) in order to analyse the strength and weakness of GP models.
- Power is well known to be influenced by air density, and this is reflected in the IEC Standard air density correction procedure. The primary objective of this research is to explore whether IEC (a traditional approach) to air density correction is the most effective when estimating power curves using a GP model. To do this definitively, SCADA datasets from turbines located in

extremely low and high-temperature regions will be used. Since temperature is the most noteworthy factor influencing air density, these datasets have air density values that are very far from the IEC Standard air density, and this is key to understanding its role in GP power curve models. The result suggests that instead of doing IEC pre-correction, adding air density directly adding into the GP model improves model accuracy and reduces the uncertainty significantly.

- In this research, the GP model for wind turbine anomaly detection is being proposed. Yaw misalignment causes significant power loss and downtime and therefore used as a case study. In order to judge GP model effectiveness, two other methods based on binning constructed have been tested and compared with GP based algorithm. In this research, we found that the GP model was able to detect the anomaly effectively with the alarm raised only 1.5 hours after the fault occurred. Hence, confirming that the GP approach provides both fast and robust fault identification.
- In this research, the impact of different turbine performance parameters (blade pitch angle and rotor speed) on GP power curve model accuracy and uncertainty presented. The inclusion of these performance parameters improves the GP model accuracy and uncertainty in which the impact of rotor speed is significant. To demonstrate this, the GP fault detection algorithm constructed with the inclusion of rotor speed and compared with GP fault detection algorithm that does not have rotor speed inclusion. The comparative analysis suggests that GP fault detection algorithm with the inclusion of rotor speed able to detect the failures after 50 minutes of the first sign of failures without any false positives while GP fault detection algorithm without the inclusion of rotor speed took approximately 1.5 hrs. This proof that the inclusion of rotor speed improves the GP model capability to detect wind turbines failures.

Finally, chapter 9 summarises the key points of each chapter along with the conclusion. In the end, future work associated with this thesis research work in this chapter.

Overall, this thesis research is of great importance for the academic and industry related to wind turbine SCADA based condition and or performance monitoring and how GP models useful in solving an issue related to condition monitoring. This gives an effective way to use GP for early failure detection, condition monitoring, performance monitoring and optimise power performance of wind turbine using extensive SCADA datasets.

1.4 Thesis overview

This thesis is organised to present the progress made and results obtained throughout the PhD study. There are 9 chapters in total, each of which will be briefly covered below.

Chapter 1 briefly summarises the research context and motivation, lists the aim and objectives and outlines scientific made by this research.

Chapter 2 describes the background and challenges associated with wind turbine condition monitoring. The state-of-the-art condition monitoring techniques researched from published literature that are commonly used for wind turbine condition monitoring are presented. Here, particular attention is paid to the nonparametric techniques based on SCADA data analysis.

Chapter 3 introduces and describes the Gaussian Process (GP) method, with a focus on regression and outlines how it can be used in the context of wind turbine condition monitoring. A comparative study of popularly used covariance functions is presented in order to assess the most suitable covariance function for GP modelling. Strengths and weaknesses associated with GP modelling are discussed briefly.

Chapter 4 proposes a comparison of the available range of statistical and machine learning based analysis methods and assesses which of these are most effective

at extracting wind turbine component condition from wind farm SCADA data. Advantages and disadvantages of the different approaches are identified.

Chapter 5 uses the method described in chapter 3 to construct wind turbine operational curves based on GP algorithms. The operational curves are constructed based on important turbine variables that can be useful in identifying faults that affect the power production of wind turbines. Also, the operational significance of SCADA data is discussed. Furthermore, GP operational curves are compared with binned operational curves with individual bin probability distributions to identify operational anomalies.

Chapter 6 outlines the importance of air density on wind turbine power curve accuracy and uncertainty. The core objective of this chapter is to explore whether the IEC approach to air density correction gives the most accurate result in case of Gaussian Process power curve or not. To do so, four possible air density approaches are proposed and analysed with the aid of SCADA datasets that manifest significant density variation from the Standard value.

Chapter 7 extends the methodology described in chapters 3 and 5 to construct a fault detection method based on a GP algorithm which then compared with two other algorithms based on binning. The proposed methods have been applied to the variable pitch wind turbine system.

Chapter 8 analyses the impact of two additional wind turbine performance parameters (blade pitch angle and rotor speed) on improving GP model accuracy and uncertainty. Based on this analysis, the most influential turbine parameter is selected and tested with GP fault detection algorithm to determine its performance in terms of capability to detect in advance (and by how much) signs of failures, and its false positive rate. This is explained with yaw misalignment (a case study) and then compared with the already constructed GP fault detection algorithm of chapter 8.

Finally, Chapter 9 concludes the research work presented in this thesis and highlights some suggestion for future study in this area.

1.5 Research Collaborations

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 642108 as part of the Advanced Wind Energy Systems Operation and Maintenance Expertise (AWESOME) Consortium [3]. The research of this thesis includes the joint collaboration of following academic and industrial partners,

- Industrial Secondment - Wood Group (previously known as Sgurr Energy), Glasgow. United Kingdom
- Academic Secondment - University of Castilla–La Mancha, Spain.

These two secondments played an essential role in developing a holistic wind turbine SCADA based condition monitoring system based on GP.

1.6. Research Publications

A number of journal publications and conference papers (including peer-reviewed) have been published during the period of this project; they are listed below with the most recent publications presented first.

Journals

1. Ravi Kumar Pandit, David Infield and Athanasios Kolios. Comparison of advanced nonparametric models for Wind Turbine Power Curves. Submitted to **Renewable Power Generation**, IET, 2019. doi: **10.1049/iet-rpg.2018.5728**.
2. Ravi Kumar Pandit and David Infield. Comparative analysis of Gaussian Process power curve models based on different stationary covariance functions for the purpose of improving model accuracy. **Renewable Energy**, **Elsevier,** 2019. vol. 140, September 2019, pp. 190-202. doi: 10.1016/j.renene.2019.03.047.

3. Ravi Kumar Pandit, David Infield. Incorporating air density into a Gaussian Process wind turbine power curve model for improving fitting accuracy. **Wind Energy, Wiley publisher.** 2019;22:302-315.doi: 10.1002/we.2285.
4. Ravi Kumar Pandit, David Infield. Comparative assessments of binned and support vector regression- based blade pitch curve of a wind turbine for the purpose of condition monitoring. **International Journal of Energy and Environmental Engineering (IJEE), Springer**, 2018. Online ISSN**:** 2251-6832. doi: https://doi.org/10.1007/s40095-018-0287-3.
5. Ravi Kumar Pandit, David Infield. Comparative analysis of binning and Gaussian Process based blade pitch angle curve of a wind turbine for the purpose of condition monitoring. IOP Conf. Series: **Journal of Physics**: Conf. Series 1102 (2018) 012037.doi :10.1088/1742-6596/1102/1/012037.
6. Ravi Kumar Pandit, David Infield. SCADA-based wind turbine anomaly detection using Gaussian Process models for wind turbine condition monitoring purposes. **IET Renewable Power Generation**, vol. 12, no. 11, pp. 1249-1255, 20 8 2018 .doi: 10.1049/iet-rpg.2018.0156.
7. Ravi Kumar Pandit, David Infield. Performance assessment of a wind turbine using SCADA based Gaussian Process model. **International Journal of Prognostics and Health Management (IJPHM),** vol. 9, no. 023, pp.8. doi: https://www.phmsociety.org/node/2492.
8. Ravi Kumar Pandit, David Infield. Gaussian Process Operational Curves for Wind Turbines Condition Monitoring. **Energies,** vol 11, no. 7: 1631, 2018. doi: 10.3390/en11071631.
9. Ravi Kumar Pandit, David Infield. Comparison of binned and Gaussian Process based power curves for condition monitoring purposes. **Journal of Maintenance Engineering** (vol: 2), University of Manchester, 2017. ISBN no: 978-1-912505-25-8.

Conference proceedings

1. Ravi Kumar Pandit, David Infield. 'SCADA based nonparametric models for condition monitoring of a wind turbine' on **The 7th International Conference on Renewable Power Generation.** (Status: accepted and in press).
2. Ravi Kumar Pandit, David Infield. QQ plot for assessment of Gaussian Process predicted power curve error distribution function. **9th European Workshop on Structural Health Monitoring (EWSHM 2018)**, July 10-13, 2018 in Northampton, UK.
3. Ravi Kumar Pandit, David Infield. Power curve modeling using support vector machine and its accuracy dependence on kernel scale. **Fifteenth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies (CM2018/MFPT2018)**, Nottingham, 2018 (Status: accepted and in press).
4. Ravi Kumar Pandit, David Infield. Comparative study of binning and Gaussian Process based rotor curves of a wind turbine for the purpose of condition monitoring. 3rd **International Conference on Offshore Renewable Energy** (CORE 2018), Glasgow (Status: accepted and in press).
5. Ravi Kumar Pandit, David Infield. Comparative analysis of binning and support vector regression for wind turbine rotor speed based power curve use in condition monitoring. **53rd International Universities Power Engineering Conference (UPEC)**, Glasgow, 2018, pp. 1-6. doi:10.1109/UPEC.2018.8542057.
6. Ravi Kumar Pandit, David Infield. Using Gaussian process theory for wind turbine power curve analysis with emphasis on the confidence intervals**. 6th International Conference on Clean Electrical Power (ICCEP)**, Santa Margherita Ligure, 2017, pp. 744-749. doi: 10.1109/ICCEP.2017.8004774.

1.7 Chapter references

1. Article on ‘A guide to UK offshore wind operations and maintenance 2013’ published on Scottish-enterprise, 08 November 2013.
2. Van Horenbeek, A., Van Ostaeyen, J., Duflou, J. R., Pintelon, L., Quantifying the added value of an imperfectly performing condition monitoring system–application to a wind turbine gearbox, Reliability Engineering & System Safety, vol. 111, pp. 45-57, 2013.
3. Advanced Wind Energy Systems Operation and Maintenance Expertise (AWESOME). Available online at http://awesome-h2020.eu/. Accessed 30th August 2018.

Chapter 2

Wind Turbine condition monitoring, trends, challenges, and state-of-the-art techniques

Published trends show the wind energy to be one of the fastest growing renewable energy sources in the world today, and due to the recent technological advancement, wind turbine overall costs have decreased substantially. However, as the demand for wind energy continues to grow, there is a constant need to reduce operation and maintenance (O&M) costs, minimise downtime and improve the reliability of turbines. This topic is, therefore, an emerging research interest. Offshore wind farm O&M is much costlier and more demanding than onshore. Condition monitoring is a process commonly used for early detection of faults, to minimise downtime and increase the reliability of wind turbines and thus ultimately reduce the O&M cost.

In this chapter, a general overview of wind turbine condition monitoring methods (both traditional and SCADA based) and techniques with a focus on offshore wind turbines will be presented. A brief overview of new trends and future challenges in context with wind turbine condition monitoring is also provided.

2.1 Condition monitoring definition, strategies, and advantages

At this point, it is beneficial to define what is meant by condition monitoring, maintenance strategies and describe how these relate to other techniques used in the O&M of wind turbines, such as alarm and shut down systems or techniques for failure and problem investigation.

By definition [4], condition monitoring is a process of monitoring the performance of a machine, in order to identify potential changes which are indicative of a developing fault before machine reaches a stage where catastrophic damage occurs. The O&M activities are a leading component cost within the total expenditure of a wind farm project and reducing these costs is a target for current research and business models. Traditional maintenance techniques result in significant downtime and a premature replacement of components.

Robust, cost-effective condition monitoring techniques can improve machine efficiency and reduce machine downtime. The maintenance strategies used in the wind industry today can be divided into three categories [4, 5] as follows:

- Corrective maintenance (run-to-failure): This maintenance strategy performed after a failure/fault has occurred. This causes significant revenue loss and machine downtime; it, therefore, should be avoided wherever possible.
- Scheduled (preventive) maintenance: As the name suggests, this is performed at a given time period to prevent the occurrence of failures and therefore bypass or at least reduce unscheduled maintenance and associated costly downtime.
- Predictive (condition-based) maintenance: A predictive (condition-based) maintenance strategy is based on continuous condition monitoring of the machine while in operation in order to maximise power production and prevent unexpected catastrophic failures.

The cost associated with these three maintenance strategies concerning a number of failures is shown in Figure 2.1 from [4]. The prevention cost is high while repair cost is low in case of scheduled maintenance since many critical failures will not happen. Therefore, scheduled maintenance is effective in minimising the number of failures, but it is expensive. With a corrective maintenance strategy, a large number of failures will occur, and that leads to high repair costs but low prevention cost. A suitable combination of scheduled and corrective maintenance strategies can improve the reliability, downtime, and O&M as shown in Figure 2.4.

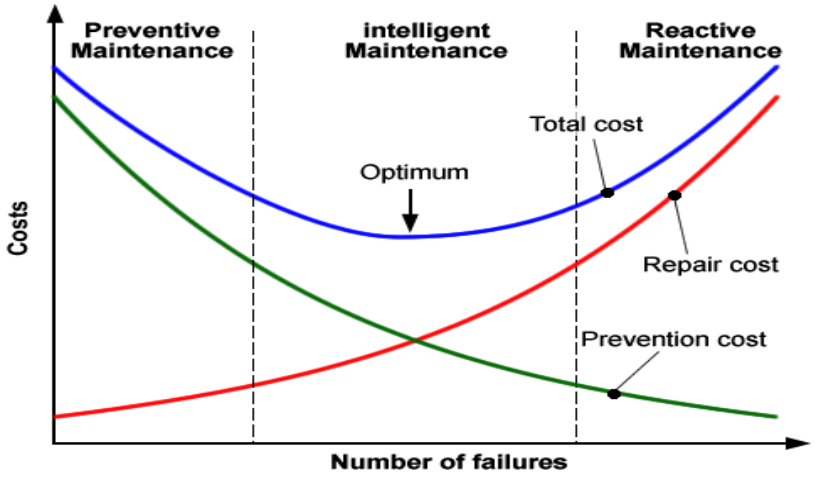


Figure 2.1: Traditional maintenance strategies cost relationship [4].

Unscheduled maintenance costs are a significant contributor to the overall wind turbine O&M cost and are generally divided into direct and indirect costs. The direct costs include the labour cost, component costs, and equipment cost (required for repair or replacement) in addition to the cost of any consumables. The revenue lost due to turbine downtime is under indirect cost, and it varies depending upon the total repair downtime, including acknowledgement, access, diagnosis, labour, and parts mobilisation, and the replacement or repair activity. Lost generation during the downtime depends on the wind resource during the time period. Improving the reliability of a turbine that is driven by the objective of minimising the O&M cost using more reliable system configurations and

components and strategies will have an impact on the levelized cost of electricity (LCOE), [6]. LCOE is used to evaluate the life-cycle costs of generation of an energy project accounting for the installed capital cost, the annual operating expenses, as well as the annual energy production. This metric allows calculation of the cost per unit of electricity generated, expressed in £/MWh.

2.2 Wind Turbine Failure Type, Rate and Patterns

Wind energy is a costly business due to the involvement of expensive electrical and mechanical components in which tower (26.3%) and rotor blades (22.2%) are the two most expensive components whose combine cost is almost equal to the half of the total cost, [7]. The gearbox (12.9%) is the 3rd most expensive component and, compared to other sub-assemblies, is prone to have high downtime per failure, [8,9]. This is because of the complex repair and maintenance procedures involved, especially for offshore wind turbines, [10].

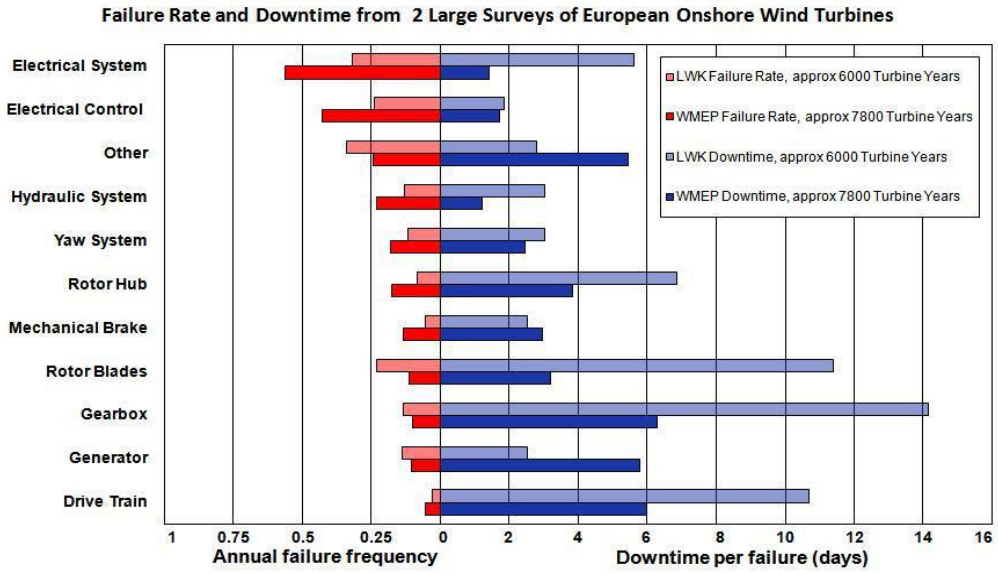


Figure 2.2: Wind turbine components failure rates and the corresponding downtime per failure for two surveys of European onshore wind turbines from [11].

The failure rate for different turbine components and its corresponding downtime per failure are labelled in Figure 2.2. It has been found that the gearbox plays the

most problematic role in both surveys regarding downtime, which can be up to 2 weeks, followed by generator and turbine blade, even though its failure rate is not as high as the electrical system and electronic control.

A wind turbine is a complex system in which specific components have a high risk of failure and failure rate varies concerning scale and type. For example, the authors of [8,12] carried out a failure analysis for onshore wind turbines, and the result suggests the general trend of an increasing failure rate with turbine size. Furthermore, Haln et al. [13] presented a survey of 1500 turbines over 15 years and found that five component groups, *i.e.*, electrical system, control system, hydraulic system, sensors, and rotor blades, are responsible for 67% of failures in turbines, as shown in Figure 2.3.

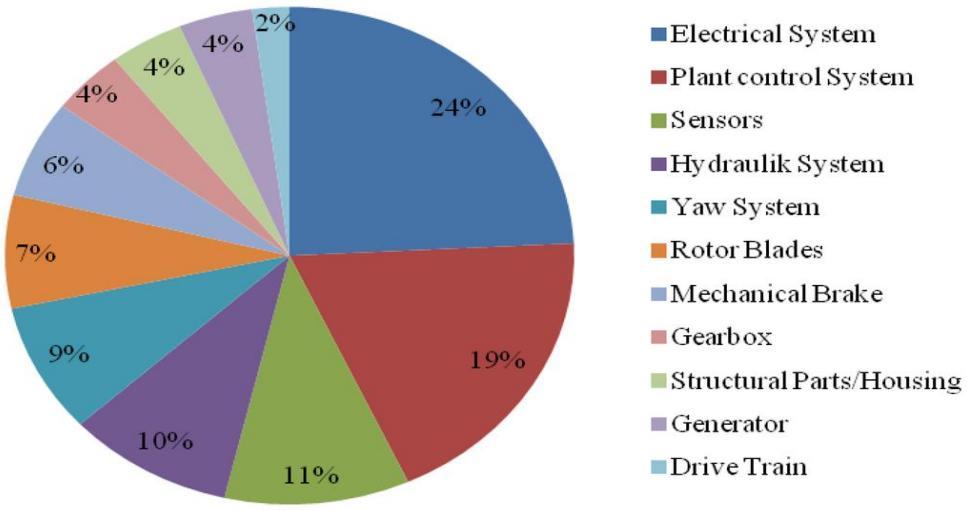


Figure 2.3: Share of the main components of the total number of failures from [13].

The machine unavailability due to failures causes revenue loss and power production loss, and there is a need to classify the failures in terms of downtime. Author of [9] classified the failures into major and minor failures. A frequently occurred failure with downtime shorter than 1 day is called a minor failure while if the failure frequency is less with downtime longer than 1 day then it is called a major failure. Furthermore, it has been found that the minor and major failures account for 75% and 25% of the onshore wind turbine failure rates respectively,

leading to a respective downtime percentage of 5% and 95%, [9]. Due to the logistics and transport issues, the downtime for both failure types increases in offshore wind turbines and turbines availability correspondingly declines. According to reference [14], the averaged technical availability for Barrow offshore wind farm in its first year of operation was as poor as 67.4%. These challenges cause expensive unscheduled maintenance, and therefore cost-effective condition monitoring can play an essential role in optimising power performance, minimising downtime, and improving the offshore turbines availability. Maintenance strategies such as those proposed in references [15], [16] and [17], can be employed to ensure the maximal turbine availability and optimal management of the available equipment taking account of the weather and the sea state.

2.3 O&M cost associated with Offshore Wind Turbines

The initial cost associated with wind energy is high due to the involvement of complex electrical as well as mechanical systems, and it is divided into two main parts, capital expenditure and operational expenditure (later referred to as CAPEX and OPEX). The CAPEX covers the cost related to wind turbine (e.g., nacelle module, tower module and rotor module), balance of plant (BOP) (e.g., assembly & installation, plant commissioning), financial (e.g. insurance during construction, project contingency budget), and land payments costs, and relatively easy to calculate while the OPEX cost related to unscheduled maintenance, repair, and spare components and difficult to calculate. In general, the O&M cost may vary with the types and size of wind turbines. The O&M costs are an essential component of the overall cost of wind energy and can vary substantially among projects and are generally costlier in case of offshore turbines. Increased competition in the wind energy sector has elevated the importance of advances in O&M in order to reduce associated costs and to ensure that turbines perform at or above expectations, especially in the case of offshore.

An offshore wind turbine needs special attention with regards to O&M because of its low reliability, remote location. They represent 25% of their lifetime costs for onshore wind farms, and 35% for offshore plants [18, 19], reducing these costs while improving availability and reliability is the current need for current O&M strategies.

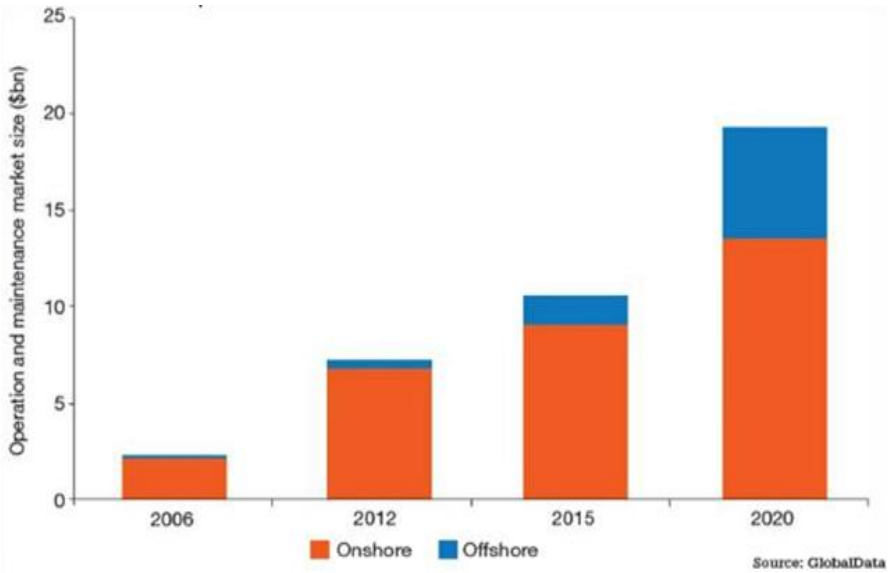


Figure 2.4: Global O&M market [21]

Several research projects on wind turbine reliability such as [20], have adopted the "bathtub curve" concept in which a constant failure rate in time during the useful lifetime of the wind turbine is assumed. This assumption may not hold true because a newly installed wind turbine needs less O&M, but as it gets older, the role of condition monitoring will be significant to increase the lifespan and efficiency of a wind turbine and offers a new business model to reduce the O&M cost. The O&M global O&M report showed the O&M market size for onshore and offshore and presented in Figure 2.4.

Good understanding of turbine parameters such as wind speed, wind shear, and pitch angle may show some significant indication of anomalous behaviour and help to predict the particular type of failure [21]. Due to the rapid increase in offshore wind turbine installation in the UK, the O&M market (Figure 2.5)

expected to grow significantly, and costs associated to O&M are substantial and can be very large compared to other costs.

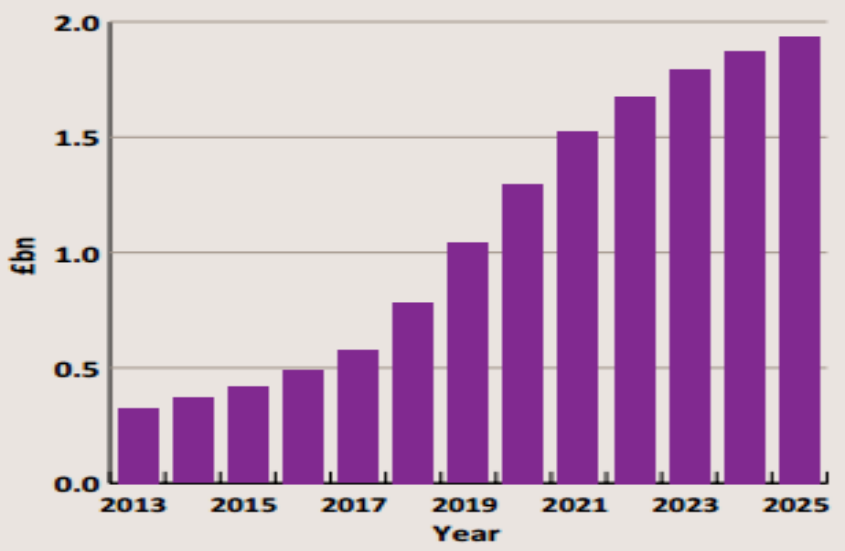


Figure 2.5: UK offshore wind O&M cost over the year's [21]

2.4 Wind turbine condition monitoring techniques

Condition monitoring is an O&M tool that helps wind turbine operators to maintain and monitor the conditions of machine components and thus better plan maintenance. It aims to make an early prediction of failures before it reaches its catastrophic stage and allow sufficient time to carry out a maintenance task. Swedish Standard SS-EN 13306 [22] defined the monitoring as an activity performed either manually or automatically with a sole objective to measure the actual state of the item. Significant effort has been made in wind turbine condition monitoring in improving turbine availability and minimising downtime. A robust condition monitoring approaches traditionally have the capability to detect the abnormal state (with locations and severity) caused by system underperformance. This section provides an overview of the methods that are commonly used by wind operators for turbine condition monitoring and can be classified into two categories; i) System physics-based techniques, and ii) SCADA based

techniques. These are introduced and described in upcoming sections, with particular attention to SCADA data-based modelling.

2.4.1 System Physics-based techniques

In system based techniques, detailed knowledge of the engineering science of the relevant system, or a sub-assembly require in order to develop an accurate condition monitoring approach for wind turbines in which exact anomaly signature needed [23]. Physics-based approaches are based on mathematical equations that signify the physics of the relevant component to assess its current and future health state and thus its performance highly influenced by the mathematical models and its capability to accurately represent the failures and deterioration phenomena [24]. For example, in references [25,26], physics-based techniques proposed and applied to detect gearbox failures. In this technique, degradation models are used to estimate long-term behaviour, and this is the main advantage associated with this technique [27]. However, in a specific condition, system based technique challenging to obtain [26].

The knowledge of the physics-based techniques is extended here to include methods based on physical system information and measurements that are obtained using additional sensors and equipment. As per [28], acoustic emission (AE), oil debris analysis, vibration signal analysis, and optical fibre (or blade strain measurement) could all play significant roles in wind turbine condition monitoring and discussed as follows.

2.4.1.1 Acoustic emission (AE)

Acoustic Emission (AE) mostly used the technique for monitoring the rotor blades on which piezoelectric sensors are used to record the release of stored elastic energy during deformation and cracking. The energy released is in the form of high energy waves which are beyond the audible range and can be categorised by their amplitude into the type of damage occurring.

The AE commonly used for fault detection in gearboxes, bearings, shafts, and blades, and its advantages include high signal-to-noise ratio (SNR) and large

frequency range. Moreover, compared to classical vibration-based methods, AE is considered to be more robust for the low-speed operation of turbines as suggested by [29]. Author of [30] found that the AE technique strongly related to vibration technique and many research was carried out on the combined use of AE and vibration techniques. For example, the authors of [31] presented the results of a combined AE and vibration monitoring for condition monitoring of turbine gearbox and generator shaft. The Risø DTU National Laboratory for Sustainable Energy [32] used the AE technique based on strain measurement to detect and locate small laminate flaws in turbine blades. AE signals are used in the past to detect the possible blade fatigue in critical areas, for example, the blade root, and briefly described in [33]. However, AE is an expensive technique and requires a very high sampling rate. Another disadvantage of AE is the attenuation of the signal during propagation which means to get an accurate result; AE sensor needs to be located as close to its source as possible which makes practical constraint in applying AE to certain wind machines [34].

2.4.1.2 Oil debris analysis

Oil debris analysis technique is an effective method for the early detection and locating the damage in bearing and gear elements in gearboxes [35] and past, it is combined with vibration analysis technique to give better accuracy in identifying gearbox failures [36]. Oil debris analysis technique used for two objectives; a) to maintain the quality of lubricant oil against the contamination by parts, moist, and b) to indicate and locate the developing faults in the mechanical components of the gearbox. The gearbox lubricant oil if have debris then it likely an indication for wear or damage of gearbox components, where a particle of different shape and material signify the different types of damage along with its location.

The oil debris analysis techniques can be classified into two categories: a) online analysis (provide instantaneous feedback of condition of the machine), and b) offline analysis (where data collected at regular intervals with sample collection

for machine condition monitoring purposes) [37]. The online analysis is a popular technique since it improves the reliability and accuracy of the analysis where failures developed rapidly or have limited accessibility [38].

Although, an oil debris analysis only widely used technique for detecting cracks in the internal part of the gearbox, however, suffers from two disadvantages. First, the equipment used for online oil analysis is expensive, and second, it cannot identify the failures outside the gearbox. Because of these reasons, offline oil analysis is often used [30, 39].

2.4.1.3 Vibration signal analysis

The vibration analysis considered to be the most widely used condition monitoring technique for rotating equipment and started finding application in the blade and tower monitoring [30]. In vibration analysis technique, accelerometers are used to measure the forces being applied to the component, and these are changes over time with frequency. It is used to monitor gearbox, bearing and shaft components of turbines but subject to sensor failures and has limited application for low-speed rotation [40]. Furthermore, it may not be ideal for all wind turbine types and faults [41]. The turbines condition monitoring techniques with vibration signal analysis are Standardised in ISO10816 [42], which define the positioning and use of sensors. Figure 2.6 describes the principle of vibration signal analysis technique and show how the time waveform of a vibration signal is decomposed into its spectral components and how characteristics frequencies in the resulting spectrum can be related to machine components.

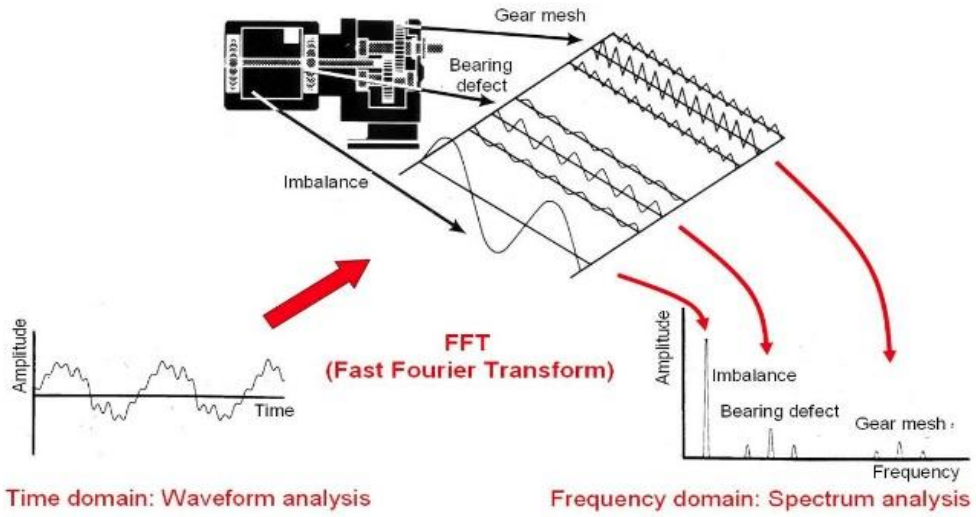


Figure 2.6: Principle of vibration signal analysis condition monitoring technique [43] adopted from SKF Vibration Diagnostics Guide (2000) [44]

The author of [36] present a report on extensive overview of vibration signal analysis techniques in context to wind turbine drivetrain condition monitoring including the Fast Fourier Transform, cepstrum processing (uses inverse Fourier Transform), bearing envelope analysis, and wavelet transforms, which are competent in catching specific frequency components under varying rotor speed due to their inherent ability to provide better frequency resolution at low frequencies and better time resolution at high frequencies [45]. Furthermore, the time series analysis of the vibration signals is used for anomaly detection, for example, in reference [10], the development of an impending gearbox bearing failure is indicated by an increasing trend of the time series representing the enveloped high-speed shaft axial vibration amplitude signal.

2.4.1.4 Blade strain measurement

Due to the force applied, the material experiences deformation. This change is deformation defined by strain. A device called strain gauges is used in strain measurement, and such measurement predominantly applied for load monitoring and lifetime testing for rotor blades; also used for damage detection [46,47]. Strain gauges placed randomly on the critical parts of the blade, and the finite

element method is used to process the captured data [48]. The distribution varies according to the number of transducers. In reference [49], it has been pointed out that in order to detect small damages, a large number of sensor required. In the long term, strain gauges are not robust, and hence there is a need for a more robust sensor [50] for example, Kreuzer [51], Bang *et al.* [48] and Schroeder *et al.* [46] explored the development of a high-speed-fibre Bragg-grating-based sensor array system for strain-based deflection shape estimation of wind turbine structures.

2.4.2 SCADA based Wind turbine condition monitoring techniques

Compared to system physics techniques (e.g., vibration signal analysis, strain gauge measurement, acoustic emission and so on) that require in-depth engineering knowledge, SCADA based modelling is relatively easy to construct and generalise. SCADA based condition monitoring is a cost-effective approach in which already collected data of turbines are used to detect the degradation of component health and identify the associated risk of failures.

2.4.2.1 *k*-nearest neighbor's regression (*k*-NN)

k-nearest neighbour regression (or k-NN) is a non-parametric machine learning approach that uses the average of k closest observation in training dataset to give an estimate. The k-NN algorithm can be used for classification and regression problems [52, 53]. k-NN has been widely applied to forecasting problems such as wind speed forecasting [54], electricity price forecasting [55], and solar power forecasting [56]. One of the unique features of k-NN is that instead of depending on data pattern, it compares the new samples to instances which are saved in the training phase [57]. A new value is given which is equal to the average of the value of its *k* closest neighbors. To calculate the distance between classification samples and train samples, the Euclidean distance used although some researchers also used Manhattan distance, correlation distance. Author of [58] compared these distances and found that the accuracy of Euclidean distance is

96% while Manhattan distance is 95.9% and correlation distance accuracy stands at 95.1%; suggest that all distance gives almost same accuracy results. It should be noted that the accuracy of the k-NN model significantly depends on k value. For example, if the k value is too small then it fragile to the noise and if k is too large, then it overfit the data points. Therefore, appropriate choice of k value is the key of the robust k-NN algorithms, for instance, the author of [59] found that k =250 is being suitble for wind turbine power curve modelling based on k-NN. The k-NN is a lazy learning approach where the function is only approximated locally, and all computation is deferred until classification.

The k-NN is an instanced-based method hence its training time is very small and hence more efficient in anomaly detection but it is computationally intensive and processing time depends on the size of the training data and on the parameter k.

2.4.2.2 Nonlinear state estimation technique (NSET)

The nonlinear state estimation technique (NSET) is a non-parametric method, proposed by Singer [60] and currently popularly used in nuclear power plants such as for sensor calibration, lifespan predication of electric components and other research into component ageing [61]. NSET is also gaining popularity in condition monitoring and or other applications of a wind turbine. In [62], NSET has been used to model turbine tower vibration for blade angle symmetry detection. The blade angle symmetry is a common fault which leads to critical fatigue damage hence early detection help to limit fatigue damage and improve the turbine efficiency. Also, this paper concluded that condition monitoring would improve significantly if the information from the vibration signals is accompanied by other relevant SCADA data, for example, power performance, wind speed and rotor loads. Furthermore, in [63], NSET method used for wind turbine condition monitoring for a SCADA based system.

NSET and Artificial Neural Network (ANN) both are the data-driven approaches. However, there is a considerable difference between these two methods as described below,

- ANN model structure is complex and difficult to determine. With simple ANN structure, modelling activity would be weak while with complex ANN structure leads to an over-fitting problem. However pre-determined structure does not require for NSET due to its non-parametric properties as demonstrated by [64].
- In order to train the model, ANN uses historical data. It transfers the information from training data into the weights. This weight us constant in nature and has no meaning while in case of NSET, weights describe the similarity between the new input vector and vectors already in the memory matrix.

2.4.2.3 Clustering

Clustering is a parametric, machine learning algorithm where a set of n observations divided into g groups, called clusters, so that members of the same group are more alike than members of different groups. A self-organizing feature map (SOFM) is a good example of the unsupervised clustering technique that forms neurons located on a regular grid, usually in one or two dimensions [65]. Catmull [66] and Kim et al. [67] were the first to use an artificial neural network (ANN) SOFM technique to wind turbine SCADA data where clusters are used to rearranging neurons on a regular grid during the training process in a way that neighbouring neurons denote similar input data. To visualize the clustering, a unified distance matrix can be used and in combination with projections of parameters, interpretation of the clustering can be started. To detect the abnormality, authors of [66] train the model and proposed the calculation of the distance between new input data and the best matching neuron. In this study, generator failures are used as a case study and result demonstrate constructed

algorithm capabilities to detect failures early using the turbines data. Furthermore, the author of [67] used a training dataset, which included failures and then they were able to select successive turbine failures to interrelated clusters.

The author of [68] used ANN SOFM to improve the Reliability and Maintainability of Wind Turbines in which based on clustering results, O&M strategies are planned for a group of wind turbines in contrast to doing the same work for the individual turbine. Zhang and Kusiak [69] proposed three SCADA (of 10-minute sampling interval) based approach for identifying abnormal vibration in the time domain. The modified k-means clustering algorithm used in first vibration monitoring model. The k-means algorithm grouped data into clusters by examining their similarity and Euclidean distance used to calculate the distance between data and cluster centroids. The clusters were then designate as the normal or abnormal status of the turbine based on error reports. The control charts are used to construct the models for monitoring of turbine vibration and extended model applications in detecting abnormal drive train and tower vibration of a wind turbine.

2.4.2.4 Artificial Neural Networks (ANNs)

Artificial neural network (ANN) is a non-parametric model inspired by biological nervous system emulate the natural intelligence of the human brain [70] and extensively used in wind turbine condition monitoring. The basic structure for ANN modelling includes one input layer, a variable of hidden layers, and one output layer. Each layer includes a large and different number of neurons, which are connected by all inputs or neuron output from the preceding layer in which each neuron consists of a nonlinear transfer function to merge the inputs and an activation function deciding if the output is generated [71,72]. The primary learning methodology of ANN is all about changing the input weight [73].

Authors of [74] presented a GP-ANN based technique for calculating the power curve of any wind turbine that depends merely on a minimal group of input variables such as hub height wind speed, and wind direction from the turbine nacelle. The author used GP for filtering and or pre-processing the SCADA data of wind turbine, and after filtration, the model trains ANN with a different number of neurons and select the final configuration considering the RMSE and the annually estimated energy error, both are calculated on the validation dataset. They carried out a comparative analysis of this proposed technique with other parametric and non-parametric methods and resulted that the GP-filtering significantly improves the performance of all methods. Overall, the combination of the GP-filtering with ANN model offers an automatic and accurate method for calculating the power curve, with a significant improvement (25%) in RMSE of predicted power as compared to the Standard method [74].

F. Pelletier et al. [75], the multistage approach presented for power curve estimation where air density and wind speed are used as input in order to obtained normalised power output in the first stage then wind speed data turbulence intensity being used to train the second ANN stage. This method produced better results compare to parametric, discrete and non-parametric models.

Pramod Bangalore et al. [76] present a self-evolving maintenance scheduler (SEMS) framework for maintenance management of wind turbines and proposes an ANN-based condition monitoring approach based on SCADA data. The majority of failures in the gearbox originated from the gearbox bearings and considered to be a significant contributor toward downtime, and early detection of it improves the machine downtime and hence used by the author as a case study. SCADA data from onshore wind turbines (of rated 2 MW) located in the south of Sweden used by authors. The result suggests that the ANN-based condition monitoring technique is capable of indicating an anomaly in advance with high accuracy and compared with previously publishing similar techniques, the main strength of the proposed model is its simplicity in terms of application.

Furthermore, Brandão et al. [77, 78] proposed FSRC ANN technique to the gearbox and generator fault detection in Portuguese wind farm (have 13 turbines with 2 MW rated power) and a USA wind farm (have 69 turbines with 1.5 MW rated power) but the author did not provide more details about the setting.

Zhang and Wang [79] used ARX ANN approach to the main shaft rear bearing temperature in direct-drive wind turbines. Approximately based on one year of SCADA data from two 3 MW wind turbines in a 17 wind turbine farm, a failure in one turbine was detected three-month advance with this model where power, nacelle temperature was used as output while wind speed taken as inputs. They set the anomaly threshold to 1.5°C for the residuals and was verified with the second turbine normal operation. Furthermore, reference [80] provides the detailed and extended application of ANN model for wind turbine condition monitoring.

ANN model is robust in the nonlinear statistical identification, but special care needs to be given for relationship learning during model construction and training stage. Issues like local minima in gradient descent algorithm and extrapolation limitations in model estimation have already gained awareness for different ANN approach applications. Author of [81] studied the local minima issue of neural networks in using the backpropagation algorithm and proposes some sufficient conditions for robust solutions. Reference [82] presented an insight into the ANN model structure and explained its limited extrapolation capability.

2.4.2.5 Fuzzy system

In recent years, the application of fuzzy system on wind turbine condition monitoring has increased significantly. A typical fuzzy system inspired by fuzzy logic which is an extension of multivalued logic based on if-then rules (i.e., the degree of truth instead of Boolean logic (true/false) [83]. The fuzzy logic is a convenient way to map an input space to an output space and input mapping to a fuzzy value is defined by the membership functions (MF).

Meik Schlechtingen et al. [83] presented a method to monitor wind turbine SCADA data, via normal behaviour models based on fuzzy logic whose sole objective is to detect trends and patterns in SCADA data and predict possible failures. The Adaptive Neuro-Fuzzy Interference Systems (ANFIS) technique can signify the nonlinear signal relations by introducing a set of fuzzy rules and tuning the Membership Function (MF) parameters in a training stage, and they used this technique for condition monitoring purposes. They used fuzzy logic in anomaly/prediction error pattern interpretation and fund that root cause diagnosis can be implemented intuitively. This gives an opportunity for automated fault diagnosis after specific rules are achieved. Furthermore, the author builds ANN (analogue to the one described in [84]) model in which only five runs with random weight initialisations are performed. The comparative analysis of ANN and ANFIS shows that both approaches performance is similar in terms of standard deviation, but training time is smaller in ANFIS as compared to ANN. The ANN takes a long time to train the model due to its necessary trial and error procedure. However, the appropriateness of the proposed model is dependent on the availability of a broad variety of different SCADA data to set up the ANFIS models.

Li et al. [85] proposed an improved fuzzy synthetic model based on a real-time condition assessment technique of a grid-connected wind turbine generator system (WTGS) to improve the operational reliability and optimise the maintenance strategy. They proposed an improved fuzzy synthetic condition assessment method which utilises the concepts of deterioration degree, limited dynamic values and variable weight calculations of the assessment indices. The 850 kW WTGS data are used in a proposed fuzzy synthetic method for real-time condition assessments. The proposed model is then compared with a traditional fuzzy assessment method in which constant limited values and constant weights are taken up. The comparative analysis of these two methods shows that the proposed model able to predict the change in operating conditions and has a better

coherence with real operating conditions than that of a traditional fuzzy assessment technique.

Other techniques such as probabilistic, trend analysis and turbines operational curves based are significantly being used for the wind turbine condition monitoring. These techniques, in general, make condition monitoring robust and reduces the O&M cost. In upcoming chapters, these techniques state of the art in context with condition monitoring would be discussed.

2.5 Why use SCADA data?

Supervisory control and data acquisition (SCADA) system is a communication and control system used for monitoring, operation and maintenance of various systems such as wind power generation or power systems. The power industries have utilised SCADA information for over 35 years, and in wind turbine, SCADA role is to supervise basic turbine operation such as the turbine cut-in, cut-out and emergency stop [86]. A SCADA system is designed to supply data related to operational status but not necessarily the health of a wind turbine (WT).

All modern wind farms are equipped with SCADA systems that record vital operational parameters, broadly classified into controllable parameters (e.g. blade pitch angle and generator torque), and non-controllable parameters (e.g. wind speed, ambient temperature) and performance parameters (e.g. power, generator speed, rotor speed). SCADA systems typically provide 10-minute averaged signals, often with associated standard deviation and maximum and minimum values. This recorded information contains continuous time observations that can be utilised for overall turbine performance monitoring as well as play a significant role in identifying component faults, at no additional cost. Kusiak and Wenyan [87] constructed SCADA based model to detect the fault at three different levels: (1) fault and no-fault prediction; (2) fault category (severity); and (3) the specific fault prediction; results show that faults can be predicted 50-60

minutes in advance at all three levels. However, the time period is short and does not give enough time to the operator to carry out maintenance actions.

Advantages

- The traditional condition monitoring system (CMS), for example, vibration analysis and oil debris detection, require expensive sensors. The additional cost can be significant considering the number of turbines typically deployed in offshore wind farms and also costly expertise is generally required to interpret the results. In contrast, SCADA data is already in place and can add real value to the condition monitoring with little or no cost to the wind farm operator.
- With the help of SCADA data, incipient wind turbine operational faults or failures can be detected before they evolve to catastrophic failures and so that preventative maintenance or corrective action can be scheduled in time, reducing downtime and potentially preventing more extensive damage. Use of SCADA should significantly reduce operating and maintenance cost and increase reliability. The author of [73] reviewed SCADA based wind turbine condition monitoring focusing on approaches which have already proved their ability to detect anomalies in data from real turbines.
- Wind turbine SCADA data comprises more than 100 different signals, ranging from the timestamp, calculated values, set point, measurements of temperature, current, voltage, wind speed, power output, wind direction and so on. These massive volumes of SCADA data make analysis effective and can be linked with traditional CMS.

Disadvantages

- SCADA data typically record 10-minute averaged data which makes diagnosis results less accurate. However, this issue in analysis depth could be minimised

by the width of the signal that is given by SCADA data across the key components as demonstrated by [86].

- The wind turbine SCADA system does not collect all of the information necessary to conduct a full turbine health assessment [73].
- Sensor failures and data malfunction cause corrupt SCADA data and errors that make model analysis inaccurate and confusing and hence should not be used in the training stage. To obtain the best results, prior pre-processing of these data points is essential.

❖ SCADA data pre-processing

The shutdown of turbines for some reason other than anomalous operation and null entries or confusing data makes SCADA data erroneous which can affect the model accuracy and further analysis. Therefore, it is important to remove these misleading or confusing informations before using this data for further analyses. Wind industries remove these SCADA errors manually. For example, in ref. [88], authors proposed a technique based on visual inspection of the power curve and can be removed before further model construction and data analysis. However, proposed techniques accuracy suffers due to the involvement of massive SCADA data. The various statistical approach has been proposed that removes these outliers before the development of models. A probabilistic technique [89] based on the copula-based joint probability model for wind turbine power curve outlier rejection is proposed. In [90], authors used alarm logs signals recorded by the SCADA system to detect the outliers where multivariate curve techniques are being used. However, in many SCADA system, alarm logs are not being stored accurately; hence this method may not work in such cases. Nevertheless, an appropriate method for SCADA data preprocessing is important for effective model constructions.

The SCADA system records an extensive amount of wind turbines data which are generally divided into wind parameters (e.g., wind speed and wind deviations), performance parameters (e.g., power output, rotor speed, and blade pitch angle)

vibration parameters (e.g., tower acceleration and drive train acceleration) and temperature parameters (e.g., bearing temperature and gearbox temperature) [91].
Many researchers used this information to preprocessed SCADA data manually such as the criteria are outlined in [84], for example, abnormal wind speed, timestamp mismatches, out of range values, negative power values, and turbine power curtailment are used to remove misleading data. This method does not make SCADA data error free but minimizes its impact significantly. Hence, this methodology would be used for SCADA data filtration in this thesis.
For wind farm owners and operators, reducing the O&M cost and maximising the availability are the main priorities and therefore SCADA based condition monitoring can be useful in this regards where failure can be detected earlier allowing sufficient time for maintenance and minimising machine unavailability and repair costs. Numerous techniques have been successfully used SCADA data for wind turbine condition monitoring (see section 2.4) where inter-relationship between critical operational parameters have been explored for effective anomaly detection algorithm or performance monitoring.

2.6 Chapter conclusions

The importance of performance and or condition monitoring in wind turbines are briefly presented in this chapter along with current trends and challenges. Different condition monitoring techniques are briefly described in this chapter, and the selection of a particular technique depends on the nature of the available data and expert domain knowledge. Traditional condition monitoring techniques (system Physics-based) are sophisticated and requires expensive sensors. Furthermore, the cost to maintain their performance is also higher. Therefore, wind industries are shifting their focused towards cost-effective SCADA based technologies for their wind turbines performance improvements.
Numerous research has been carried out on SCADA based condition monitoring, but its commercial applications are limited because wind farm operators remain

sceptical about the benefits of SCADA based technologies. In the last decade, the involvement of the SCADA system for wind turbine applications are significant. Various parametric and nonparametric (including machine learning) approaches have been using SCADA data for wind turbine performance or condition assessment. It has been found that nonparametric techniques perform better than parametric models due to their robustness. Nonparametric techniques such as ANNs, Fuzzy system, k-NN and NSET have popularly used wind turbine condition monitoring and are well described in this chapter.

The SCADA data recorded in the system contains errors, for example, time mismatch, negative power values and need to filter out for proposed models accuracy. Many researchers proposed different techniques and are well described in section 2.5. SCADA data pre-processing described in [84] are adopted in this thesis because it is simple, effective and uses already available information without extra mathematical computations.

Importance of SCADA datasets are presented in this chapter that contains the information about the operational status of wind turbines. Therefore, SCADA based condition monitoring approach is being proposed for monitoring the continuous performance of turbines where the extensive applications of GP (a nonparametric, machine learning method) is being explored and outlines SCADA based GP models strength and weakness as compared to available nonparametric approach (including traditional methods)

2.7 Chapter references

4. Orsagh, R.F.; Lee, H.; Watson, M.; Byington, C.S.; Power, J. Advance Vibration Monitoring for Wind Turbine Health Management; Impact Technologies, LLC: Rochester, NY, USA, 2006.
5. Dhillon, B.S. Engineering Maintenance: A Modern Approach; CRC Press: Boca Raton, FL, USA, 2002.
6. IRENA (2016). The power to change: Solar and Wind cost reduction potential to 2025. Accessed on 10th June 2018.
7. IRENA (2012). Renewable Energy technologies: cost analysis series. Volume 1: power sector, issue 5/5. Accessed 26th June 2018.
8. F. Spinato, P.J. Tavner, G.J.W van Bussel, E. Koutoulakos. Reliability of wind turbine subassemblies. IET Renewable Power Generation, 3 (4) (2009), p. 387. doi: 10.1049/iet-rpg.2008.0060.
9. S. Faulstich, B. Hahn, PJ. Tavner. Wind turbine downtime and its importance for offshore deployment. Wind Energy, 14 (2010), pp. 327-337.
10. Y. Feng, Y. Qiu, C. Crabtree, H. Long, P. Tavner. Monitoring wind turbine gearboxes. Wind Energy, 16 (5) (2013), pp. 728-740.
11. Crabtree, C.J., Feng, Y., Tavner, P.J., Detecting Incipient Wind Turbine Gearbox Failure: A Signal Analysis Method for Online Condition Monitoring, Scientific Track Proceedings, European Wind Energy Conference 2010, Warsaw, Poland, 2010.
12. Tavner, P.J.; Xiang, J.; Spinato, F. Reliability analysis for wind turbines. Wind Energy 2007, *10*, 1–18.
13. Hahn, B.; Durstewitz, M.; Rohrig, K. Reliability of Wind Turbines. In Wind Energy; Springer: Berlin/Heidelberg, Germany, 2007; pp. 329–332.

14. Feng, Y., Tavner, P.J., Long, H., Early experiences with UK Round 1 offshore wind farms. Proceedings of the Institution of Civil Engineers: energy. vol. 163, (4), pp. 167-181., 2010.
15. Mahmood Shafiee and John Dalsgaard Sørensen. Maintenance optimization and inspection planning of wind energy assets: Models, methods and strategies. Reliability Engineering & System Safety, 2017. doi: 10.1016/j.ress.2017.10.025.
16. Albert H. Schrotenboer, Michiel A.J. uit het Broek, Bolor Jargalsaikhan and Kees Jan Roodbergen. Coordinating technician allocation and maintenance routing for offshore wind farms. Computers & Operations Research, vol 98, October 2018, pp. 185-197. doi: 10.1016/j.cor.2018.05.019.
17. Chandra Ade Irawan, Djamila Ouelhadj, Dylan Jones, Magnus Stålhane and Iver Bakken Sperstad. Optimisation of maintenance routing and scheduling for offshore wind farms. European Journal of Operational Research,2017, vol 256 (1), pp. 76-89 ,2017.
18. Y. Sinha, J. Steel. A progressive study into offshore wind farm maintenance optimisation using risk based failure analysis. Renewable and Sustainable Energy Reviews, 2017, vol 42, pp. 735-742. doi: 10.1016/j.rser.2014.10.087.
19. María Isabel Blanco. The economics of wind energy. Renewable and Sustainable Energy Reviews,2009, vol 13(6-7), pp. 1372-1382. doi: 10.1016/j.rser.2008.09.004.
20. Ribrant, J. & Bertling, L. M., 2007. Survey of failures in wind power systems with focus on Swedish wind power plants during 1997-2005. IEEE Transactions on Energy conversion, March, 22(1), pp. 167-173
21. A Guide to UK Offshore Wind Operations and Maintenance, by GL Garrad Hassan.
22. Maintenance Terminology; SS-EN 13306; Swedish Standards Institute: Stockholm, Sweden, 2011.

23. Hyers, R. W., McGowan, J.G., Sullivan, K.L., Manwell, J.F., Syrett, B.C., Condition Monitoring and Prognosis of Utility Scale Wind Turbines, Energy Materials, vol. 1, no. 3. pp. 187-203, Sep. 2006.
24. Adrian Cubillo, Suresh Perinpanayagam and Manuel Esperon-Miguez. A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery. Advances in Mechanical Engineering 2016, Vol. 8(8) 1–21. doi: 10.1177/1687814016664660.
25. Sepulveda M, Shek J, Thies PR, Oterkus E, Davies P, Spring M. Physics-Based Gearbox Failure Model for Multi-MW Offshore Wind Turbines. ASME. International Conference on Offshore Mechanics and Arctic Engineering, Vol 3B: Structures, Safety and Reliability (): V03BT02A011. doi:10.1115/OMAE2017-62257.
26. Christopher S. Gray, Simon J. Watson. Physics of Failure approach to wind turbine condition based maintenance. Wind Energy, 13: 395-405. doi: 10.1002/we.360.
27. An D, Kim NH and Choi JH. Options for prognostics methods: a review of data-driven and physics-based prognostics. In: Proceedings of the 54th AIAA/ASME/ASCE/ AHS/ASC structures, structural dynamics, and materials conference, Boston, MA, 8–11 April 2013, pp.1–14. Reston, VA: AIAA.
28. A. May, D. McMillan and S. Thöns. Economic analysis of condition monitoring systems for offshore wind turbine sub-systems. IET Renewable Power Generation, vol. 9, no 8, pp. 900-907,11 2015. doi: 10.1049/iet-rpg.2015.0019.
29. Ozak-Obazi Oluwaseyi Esu. Vibration-based condition monitoring of wind turbine blades. PhD thesis, 2016. Loughborough University, Leicestershire, UK.
30. Hameed, Z., Hong, Y.S., Cho, Y.M., Ahn, S.H., Song, C.K., Condition Monitoring and Fault Detection of Wind Turbines and Related Algorithms:

A Review, Renewable and Sustainable Energy Reviews, Vol.13, pp. 1-39, 2009.

31. Soua, S., Van Lieshout, P., Perera, A., Gan, T. H., Bridge, B., Determination of the combined vibrational and acoustic emission signature of a wind turbine gearbox and generator shaft in service as a pre-requisite for effective condition monitoring, Renewable Energy, vol. 51, pp. 175-18, 2013.
32. G. Marsh, "In-service Monitoring of Turbine Blades," Reinf. Plast., vol. 52, no. 5, pp. 24– 27,29, 2008.
33. Beattie, A. G., Acoustic emission monitoring of a wind turbine blade during a fatigue test, In AIAA Aerospace Sciences Meeting, REno, Nevada, January, 1997.
34. Tan, C.K.; Irving, P.; Mba, D. A comparative experimental study on the diagnostic and prognostic capabilities of acoustics emission, vibration and spectrometric oil analysis for spur gears. *Mech. Syst. Signal Process.* 2007, *21*, 208–233.
35. Barrett, M.P.; Stover, J. Understanding Oil Analysis: How It Can Improve the Reliability of Wind Turbine Gearboxes; Gear Technology: Elk Grove Village, IL, USA, 2013; pp. 104–111.
36. Sheng, S., Wind Turbine Gearbox Condition Monitoring Round Robin Study–Vibration Analysis, NREL technical report, 2012.
37. Sheng, S.; Veers, P. Wind Turbine Drivetrain Condition Monitoring—An Overview, Mechanical Failures Prevention Group: Applied Systems Health Management Conference 2011, Virginia Beach, VA, USA, 10–12 May 2011.
38. Hamilton, A.; Quail, F. Detailed state of the art review for the different on-line/in-line oil analysis techniques in context of wind turbine gearboxes. J. Tribol. **2011**, 133.
39. Hasan, I.; Rahaman, M.I. Intelligent Diagnostics and Predictive Maintenance Sensor System for Electrical Fault Diagnosis of Wind Turbine System.

Proceedings of the Global Engineering, Science and Technology Conference 2012, Dhaka, Bangladesh, 28–29 December 2012.

40. B. Lu, Y. Li, X. Wu, and Z. Yang. A review of recent advances in wind turbine condition monitoring and fault diagnosis. IEEE Power Electron. Mach. Wind Appl., pp. 1–7, Jun. 2009.
41. W. Yang, P. J. Tavner, C. J. Crabtree, and M. Wilkinson. Cost-Effective Condition Monitoring for Wind Turbines. IEEE Trans. Ind. Electron., vol. 57, no. 1, pp. 263–271, Jan. 2010. doi: 10.1109/TIE.2009.2032202.
42. Mechanical Vibration—Evaluation of Machine Vibration by Measurements on Non-Rotating Parts—Part 1: General Guidelines; ISO 10816-1:1995; International Organization for Standardization (ISO): Geneva, Switzerland, 1995.
43. J. Nilsson. Maintenance management of wind power systems—Cost effect analysis of condition monitoring systems. Master's thesis, KTH, Stockholm, Sweden, 2005–2006.
44. SKF Reliability Systems (2000)] SKF Reliability Systems (2000): Vibration Diagnostic Guide. Available at www.skf.com.
45. Watson, S. J., Xiang, B. J., Yang, W., Tavner, P. J., Crabtree, C. J. Condition monitoring of the power output of wind turbine generators using wavelets. IEEE Transactions on Energy Conversion, 25(3), pp. 715-721, 2010.
46. Schroeder K, Ecke W, Apitz J, Lembke E, Lenschow G. A fibre Bragg grating sensor system monitors operational load in a wind turbine rotor blade. Measurement Science and Technology 2006;17(5):1167e72.
47. Wernicke J, Shadden J, Kuhnt S, Byars R, Rhead P, Damaschke. Field experience of fibre optical strain sensors for providing real time load information from wind turbine blades during operation. In: Paper presented at the European wind energy conference, 22e25 November, London, UK; 2004.

48. Bang, H.-J.; Kim, H.-I.; Lee, K.-S. Measurement of strain and bending deflection of a wind turbine tower using arrayed FBG sensors. *Int. J. Precis. Eng. Manuf.* 2012, *13*, 2121–2126.
49. R. Rolfes, S. Tsiapoki and M.W.Häckell. 19 - Sensing solutions for assessing and monitoring wind turbines. Sensor Technologies for Civil Infrastructures Applications in Structural Health Monitoring, vol 2 in Woodhead Publishing Series in Electronic and Optical Materials, pp. 565-604, 2014. doi: 10.1533/9781782422433.2.565.
50. Elforjani, M.A. Condition Monitoring of Slow Speed Rotating Machinery Using Acoustic Emission Technology. Ph.D. Thesis, Cranfield University, Cranfield, UK, June 2010.
51. Kreuzer, M. Strain Measurement with Fiber Bragg Grating Sensors; S2338-1.0 e; HBM GmbH: Darmstadt, Germany, 2006.
52. T. Cover, P. Hart. Nearest neighbor pattern classification. IEEE Trans Inf Theory, 13 (1) (2006), pp. 21-27.
53. S.A. Dudani. The distance-weighted k-nearest-neighbor rule. IEEE Trans Syst, Man, Cybernet, SMC-6 (4) (1976), pp. 325-327, 10.1109/TSMC.1976.5408784.
54. M. Yesilbudak, S. Sagiroglu, I. Colak. A new approach to very short term wind speed prediction using k-nearest neighbor classification. Energy Convers Manage, 69 (2013), pp. 77-86. doi: 10.1016/j.enconman.2013.01.033.
55. A.T. Lora, J.M.R. Santos, A.G. Exposito, J.L.M. Ramos, J.C.R. Santos. Electricity market price forecasting based on weighted nearest neighbor's techniques. IEEE Trans Power Syst, 22 (3) (2007), pp. 1294-1301, 10.1109/TPWRS.2007.901670.
56. H. Long, Z. Zhang, Y. Su. Analysis of daily solar power prediction with data-driven approaches. Applied Energy, 126 (2014), pp. 29-37. doi: 10.1016/j.apenergy.2014.03.084.

57. Long Peng, Bin Jiao, Hai Liu, Ting Zhang. The Fault Diagnosis of Wind Turbine Gearbox Based on Improved KNN. doi: 10.14355.
58. Dibya Jyoti Bora and Dr. Anil Kumar Gupta. Effect of Different Distance Measures on the Performance of K-Means Algorithm: An Experimental Study in Matlab. International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014, 2501-2506.
59. Shahab Shokrzadeh. 'Wind Turbine Power Curve Modeling Using Advanced Parametric and Nonparametric Methods'. IEEE transactions on sustainable energy, vol.5, no.4, October 2014.
60. Gross, K.C.; Singer, R.M.; Wegerich, S.W.; Herzog, J.P. Application of a model-based fault detection system to nuclear plant signals. In Proceedings of 9th International Conference on Intelligent Systems Application to Power System, Seoul, Korea, 6–10 July 1997; pp. 212–218.
61. Cheng, S.F.; Pecht, M. Multivariate state estimation technique for remaining useful life prediction of electronic products. In Proceedings of AAAI Fall Symposium on Artificial Intelligence for Prognostics, Arlington, VA, USA, 9–11 November 2007; pp. 26–32.
62. Peng Guo; David Infield. Wind Turbine Tower Vibration Modeling and Monitoring by the Nonlinear State Estimation Technique (NSET). Energies 2012, 5(12), 5279-5293; doi:10.3390/en5125279.
63. Wang, Y.; Infield, D. SCADA data based nonlinear state estimation technique for wind turbine gearbox condition monitoring. In Proceedings of European Wind Energy Association Conference 2012, Copenhagen, Denmark, 16–19 April 2012; pp. 621–629.
64. Guo, P.; Infield, D.; Yang, Y. Wind turbine generator condition monitoring using temperature trend analysis. IEEE Trans. Sustain. Energy **2012**, *1*, 124–133.
65. Kohonen, T. Self-Organizing Maps, Springer Verlag, Heidelberg, 1995.

66. Catmull, S.: ‘Self-organising map based condition monitoring of wind turbines’. EWEA Annual Conf. 2011, 2011.
67. Kim, K., Parthasarathy, G., Uluyol, O., et al.: ‘Use of SCADA data for failure detection in wind turbines’. ASME 5th Int. Conf. Energy Sustain., 2011, pp. 2071–2079.
68. Zafar Hameed and Kesheng Wang. Clustering Analysis to Improve the Reliability and Maintainability of Wind Turbines with Self-Organizing Map Neural Network. International Journal of Performability Engineering, Vol. 9, No. 3, May 2013, pp. 245-260.
69. Kusiak A, Zhang Z (2010). Analysis of wind turbine vibrations based on SCADA data. J Sol Energy Eng. doi: 10.1115/1.4001461.
70. S. R. Jang, C.-T. Sun, and E. Mizutani, Neuro-Fuzzy and Soft Computing-a Computational Approach to Learning and Machine Intelligence, PHI ltd, New Delhi, India, 2003.
71. Zaher, A., McArthur, S.D.J., Infield, D.G., et al.: ‘Online wind turbine fault detection through automated SCADA data analysis’, Wind Energy, 2009, 12, (6), pp. 574–593.
72. Du, H.: ‘Data mining techniques and applications: an introduction’ (Cengage Learning EMEA, 2010).
73. J. Tautz-Weinert and S. J. Watson, "Using SCADA data for wind turbine condition monitoring – a review," in IET Renewable Power Generation, vol. 11, no. 4, pp. 382-394, 15 3 2017. doi: 10.1049/iet-rpg.2016.0248.
74. Bartolomé Manobel, Frank Sehnke et al. (2018). Wind turbine power curve modeling based on Gaussian Processes and Artificial Neural Networks. Renewable Energy Vol. 125, pp. 1015-1020. doi: 10.1016/j.renene.2018.02.081.
75. F. Pelletier, C. Masson, and A. Tahan, “Wind turbine power curve modelling using artificial neural network,” Renewable Energy, vol. 89, pp. 207–214, 2016.

76. P. Bangalore and L. B. Tjernberg, "An Artificial Neural Network Approach for Early Fault Detection of Gearbox Bearings," in IEEE Transactions on Smart Grid, vol. 6, no. 2, pp. 980-987, March 2015. doi: 10.1109/TSG.2014.2386305.
77. Brandão, R.F.M., Carvalho, J.A.B., Barbosa, F.P.M.: 'Neural networks for condition monitoring of wind turbines'. Int. Symp. Modern Electric Power System Wroclaw, Poland, 2010.
78. Brandão, R.F.M., Carvalho, J.A.B.: 'Intelligent system for fault detection in wind turbines gearbox'. PowerTech Eindhoven 2015, 2015.
79. Zhang, Z.-Y., Wang, K.-S.: 'Wind turbine fault detection based on SCADA data analysis using ANN', Adv. Manuf., 2014, 2, (1), pp. 70–78.
80. S.A. Kalogirou. Artificial neural networks in renewable energy systems applications: a review. Renew. Sustain. Energy Rev., 5 (4) (2001), pp. 373-401
81. Bianchini, M., Gori M., Maggini, M., On the Problem of Local Minima in Recurrent Neural Networks, IEEE Transactions on Neural Networks, Vol. 5, No. 2, pp. 167-177, Mar. 1994.
82. Haley, P.J., Soloway, D., Extrapolation Limitations of Multilayer Feedforward Neural Networks, International Joint Conference on Neural Network, Vol. 4, pp. 25-30, 1992.
83. Schlechtingen, M., Santos I.F., Achiche, S., Using Data-Mining Approaches for Wind Turbine Power Curve Monitoring: A Comparative Study, IEEE Transactions on Sustainable Energy, vol. 4, no. 3, pp. 671-679, July 2013.
84. M. Schlechtingen, I.F. Santos. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. Mechanical Systems and Signal Processing, 25 (5) (2011), pp. 1849-1875,

85. Li, H.; Hu, Y.; Yang, C.; Chen, Z.; Ji, H.; Zhao, B. An improved fuzzy synthetic condition assessment of a wind turbine generator system. Int. J. Electr. Power Energy Syst. 2013, 45, 468–476.
86. Tavner, P., Offshore Wind Turbines: Reliability, availability and maintenance, IET Renewable Energy Series 13, 2012.
87. A. Kusiak, W. Li. The prediction and diagnosis of wind turbine faults. Renewable Energy, 36 (2011), pp. 16-23.
88. S. Gill, B. Stephen, and S. Galloway. Wind turbine condition assessment through power curve copula modeling. IEEE Transactions on Sustainable Energy, vol. 3, no. 1, pp. 94–101, 2012.
89. Y. Wang, D. G. Infield, B. Stephen, and S. J. Galloway. Copula-based model for wind turbine power curve outlier rejection. Wind Energy, vol. 17, no. 11, pp. 1677–1688, 2014.
90. Gonzalez, E.; Reder, M.; Melero, J.J. SCADA alarms processing for wind turbine component failure detection. *J. Phys. Conf. Ser.* 2016, *753*, 072019.
91. Verma A, Kusiak A. Fault monitoring of wind turbine generator brushes: a data-mining approach. J Sol Energy Eng, 2012. doi: 10.1115/1.4005624.

Chapter 3

Introduction to Gaussian Processes models

Gaussian Processes (GPs) are very general non-linear multivariate probabilistic models based on a generic supervised machine learning approach; they are commonly used to address problems related to regression and probabilistic classification. GPs provide an attractive data analysis framework and the basis of a potentially effective automated fault detection system for the wind sector and hence recently have started to find application to wind turbine condition monitoring. Due to their simplicity and ease of construction, GP models are also applied to solving problems related to forecasting, anomaly detection and fitting related applications.

In this chapter, a brief history of GPs models is outlined and then following a presentation of GP theory and the role of covariance function; specific GP algorithms will be discussed. In GP models, training is undertaken to select the best hyperparameters which define the covariance function. Therefore, both the basis functions and their prior distribution are simultaneously specified by choice of covariance function. Depending upon the given problem, the selection of covariance function is made independently of the basis functions, and therefore performance comparison of widely used covariance functions is essential for effective GP modelling. Hence in this chapter, comparative studies of popularly used covariance functions will be presented with their respective strengths and weaknesses.

3.1 A brief history of Gaussian Processes

A Gaussian Process is a stochastic process, simple to implement, flexible and fully multivariate. It is a probabilistic model and has a general class of probability distributions on functions. GP models are named after German mathematician Carl Friedrich Gauss because a GP is based on the notion of the Gaussian distribution. GP models are not new and have been studied and used for centuries. For example, the renowned 'Wiener process' is a continuous–time stochastic process and is a particular type of GP [92].

Gaussian Processes were first applied for time series prediction. Work in this area dates back to the 1940s [93]. Starting from 1970', GPs have been used widely in the fields of geostatistics and meteorology. In statistics, GPs were used to solve slightly more general multivariate input regression problems [93]. In geostatistics, prediction using a GP is called 'kriging' and is a method of interpolation. Kriging is named after the South African mining engineer D. G. Krige by Matheron in 1973 [94]. Williams and Rasmussen [95] in 1996' first described GPR in a machine learning context which is partly inspired by the relationship shown by Neal, [96], between GP models and neural networks. There has been significant work on GPs to solve various complex issues across a wide field of application. Rasmussen and Williams [95], published a book on GPs in the context of machine learning, and most of the material covered in this chapter, and a lot more, can be found there.

The GP concept can be classified according to two distinct applications: classification and regression. This thesis is concerned with the more specific use of GP models for prediction, forecasting, and anomaly detection for wind turbines. Therefore, Gaussian Process regression (GPR) is relevant for performance and or condition monitoring purposes.

3.2 Gaussian Process versus Gaussian distribution

GPs are called a non-parametric because they assess how all the measured data are correlated instead of trying to fit the parameters of selected basis functions.

A GP is a collection of random variables having a property that the joint distribution of any of its subsets is multivariate Gaussian. Also, it should be noted that there are differences between a Gaussian Process and Gaussian distributions which are summarised in Table 3.1.

Gaussian distribution	Gaussian process
• A Gaussian distribution is a distribution over vectors.	• A Gaussian process is a distribution over functions.
• A mean and a covariance fully specify it: $x \sim \mathcal{N}(\mu, \sum)$ The position of the random variables $\boldsymbol{x_i}$ in the vector plays the role of the index.	• Mean and covariance functions fully specify it: $f \sim \mathcal{N}(m, k)$ The argument x of the random function $f(x)$ plays the role of the index.

Table 3.1: Difference between Gaussian distribution and Gaussian process [95, 97].

3.3 Gaussian Process theory

A Gaussian Process is a multivariable probabilistic based generic supervised machine learning approach that addresses problems related to regression and probabilistic classification. Notably, a GP is a stochastic process that has Gaussian-distributed finite-dimensional marginal distributions defined over the distribution of functions, i.e. each output from the GP is itself a function. GPs are widely used in probabilistic regression problems thanks to their flexibility and simplicity of construction.

A GP, in essence, is a non-parametric generalisation of a joint normal distribution for a given potentially infinite set of variables, and it is mathematically defined by its mean and covariance functions (or kernel) as given in equation (3.1),

$$Y \sim \mathcal{N}(\mu, \sum) \tag{3.1}$$

where, μ is the mean function, and $\sum$ is the covariance function that has an associated probability density function:

$$P = \frac{1}{(2\pi)^{\frac{n}{2}} |\sum|^{\frac{1}{2}}} \, exp\left\{-\frac{1}{2}(x-\mu)^T \sum^{-1}(x-\mu)\right\} \tag{3.2}$$

Where $|\sum|$ is defined as the determinant of $\sum$, n is the dimension of random input vector x , and μ is the mean of vector x. The term under the exponential i.e. $\frac{1}{2}(x-\mu)^T \sum^{-1}(x-\mu)$ is an example of a quadratic form.

In GP models, the mean function is set to zero by appropriate renormalising of the variable x. Hyperparameters that determine the covariance function characterises GP models; these have a substantial impact on the posterior model accuracy. A GP model estimates these parameters by maximising the marginal log-likelihood of the model which depends on the nature of data, see [95]. The proper optimisation of these hyperparameters ensures GP model accuracy, and particular in this work, power curve model accuracy, and is described as follows. For a given training dataset, $A = \{(x_i, y_i), i = 1, \ldots\ldots, N\}$ of n observations, calculated values of mean and variance are required in order to model the GP power curve. Here x is the input vector of dimension D , and y is the scalar output. The given set of input datasets is denoted by a $A \times n$ matrix. Suppose for a given N training data input and output pairs (X, y), we can define the relationship as $A = (X, y)$. The GP regression is used to solve the relationship between input and target values which is modelled as:

$$y_i = f(x_i) + \epsilon_i \tag{3.3}$$

The above equation is theoretically used to define the underlying function of the data modelled where x are values from the training datasets and ϵ is Gaussian white noise of variance σ_n^2 such that, $\epsilon = N(0, \sigma_n^2)$. And the prior to *y* becomes:

$$E|\mathrm{y}| = \mathrm{E}|f + \epsilon| = 0 \tag{3.4}$$

$$cov\ |\mathrm{y}| = K|X, X| + \sigma_n^2 I \tag{3.5}$$

The prior distribution is useful in providing essential information about the uncertain parameters. This prior distribution with the probability distribution is used to develop the posterior distribution, which is helpful for inference and in any decisions involving uncertain parameters [95]. To estimate the output *f*, for a new input x^*, the distribution can be written as:

$$\begin{pmatrix} y \\ f^* \end{pmatrix} \sim N\left(0, \begin{bmatrix} K(X,X) + \sigma_n^2 I & k(X,x^*) \\ k(x^*,X) & k(x^*,x^*) \end{bmatrix}\right) \tag{3.6}$$

where $k(X, x^*) = k(x^*, X)^T = [k(x_1, x^*), k(x_2, x^*), \ldots . k(x_n, x^*)]$, which is for simplicity denoted by k^*. Then, from the joint Gaussian distribution, the estimation of target values is given by:

$$\overline{f\ *} = k_*^T (K + \sigma_n^2 I)^{-1} y \tag{3.7}$$

$$Var[f\ *] = \mathrm{k}(x^*, x^*) - k_*^T k^* (K + \sigma_n^2 I)^{-1} \tag{3.8}$$

The obtained posterior variance ($Var[f\ *]$) is inversely proportional to the distance between test and training data points while the estimation of the mean ($\overline{f\ *}$) is a linear combination of the output y in which linear weights are defined as $k_*^T (K + \sigma_n^2 I)^{-1}$. The ($\overline{f\ *}$) gives the estimated values of power (in case of power curve) while $Var[f\ *]$ yields variance associated with GP models and it is useful in calculating confidence intervals (CIs). The CIs measures the uncertainty and significant on constructing fault detection algorithms and these are well described in upcoming chapters.

3.4 Advantages and disadvantages of Gaussian Process models

Gaussian Processes not only relax the assumption of linearity in the regression parameters but also give freedom where we do not need to specify a precise functional form for the relationship between predictor and response variables. The advantages of GP are summarised as follows, [95].

Advantages

- GP models, not only estimate predicted values but also come with intrinsic confidence intervals. These confidence intervals signify the uncertainty associated with the prediction and are very useful for anomaly detection which in the intended application are wind turbine faults.
- Easy to interpolate training data.
- GP models are very versatile due to the abundant choices available for covariance function for any given problem, and the selection of hyperparameters.
- Provide smooth nonlinear models.
- With a GP, the predicted values are interpolations between the training data points that conform to the chosen covariance functions, so that the nature of the interpolation is controlled by the selection of covariance function.
- GPs can be easily extended and incorporated into hierarchical Bayesian models.

Uncertainty assessment is the key to developing a fault detection algorithm based on the GP. The uncertainty in the GP expressed by its Confidence Intervals (CIs) whose robustness is directly related to covariance function and will be discussed in upcoming sections.

Disadvantages

GP theory described above suggests that posterior computation in GP regression is a trivial matter, but there are two main hurdles involved that are described as follows.

- The computation of the mean and covariance in the *n*-variate normal conditional posterior distribution for $f*$ involves matrix inversion that requires $O(n^3)$ computation. The posterior conditional distribution for a given observation mathematically is defined in [98], and there is a covariance matrix component, as given by equation (3.9), associated with the inverse matrix operation which leads to the mathematical challenge of inverting an $n \times n$ matrix (and this goes approximately with $O(n^3)$, where n is the number of data points). This computation needs to be repeated for each step with changing hyperparameters. If n is too large, then the computation of the $n \times n$ matrix becomes problematic and leads to GP model inaccuracy.

$$|Y_x = y_x \sim GP \left[\begin{array}{l} \mu(x^*) \sum(x^*, x) \sum(x, x)^{-1} \left(y_x - \mu(x)\right), \\ \sum(x^*, x^*) - \sum(x^*, x) \sum(x, x)^{-1} \sum(x, x^*) \end{array} \right] \quad (3.9)$$

- The prediction cost per test for mean $\overline{(\boldsymbol{f}*)}$, variance $(\boldsymbol{Var}[\boldsymbol{f}*])$ and $(\boldsymbol{K} + \boldsymbol{\sigma_n^2 I})$ are $\boldsymbol{O(n)}$, $\boldsymbol{O(n^2)}$ and $\boldsymbol{O(n^3)}$ respectively and these are computationally expensive especially in dealing with big datasets.

State-space representations of GPs are being introduced to address this cubic problem, [99,100]. Moreover, in [101], two parallel GP regression methods that exploit low-rank covariance matrix approximations for distributing the computational load among parallel machines are being used to solve these problems, but these still require high processing power and computational cost in dealing with large datasets. The accuracy of a GP model depends upon the quantity and quality of the data. For example, to model a wind turbine power curve, a low number of power-wind speed pairs may not give a smooth power

curve while a high number is not desirable because of the numerical challenge of calculating the GP, [102].

3.5 Gaussian Process covariance functions

Covariance functions are commonly known as positive semi-definite kernels or Mercer's kernels, [103] and considered to be the driver of the fitting quality of the models along with hyperparameters (described above). The covariance function is the decisive ingredient in Gaussian Process modelling since it encodes the assumptions about the function which we wish to learn. A covariance function describes the dependency of two variables and is the core of the GP model; it signifies the similarity between two points and hence determines closeness between two points.

The basics properties of a covariance function are:

- It must be positive semi-definite.
- It can be non-stationary.
- It can be the sum (or product or linear combination) of other covariance functions, e.g., can use a different covariance function for each unique sensor modality or data type (vector, sequence, image data).

The reliability of GP regression is dependent on what and how well we select the covariance function. The covariance matrix, *K*, gives the variance of each variable along the leading diagonal, and the off-diagonal elements measure the correlations between the different variables and are given by:

$$K = \begin{bmatrix} k_{11} & \cdots & k_{1n} \\ \vdots & \ddots & \vdots \\ k_{n1} & \cdots & k_n \end{bmatrix} \quad \text{where } k_{ij} = k(x_i, x_j)$$

K is of size $n \times n$, where n is the number of input parameters considered, and it must be symmetric and positive semidefinite i.e. $\sum_{ij} = \sum_{ji}$.

In general, covariance functions are classified into two categories; i) stationary and ii) nonstationary. A stationary covariance function is one that only depends

on the relative position of its two inputs, and not on their absolute location in time. The stationary covariance function requires less parameter, in particular as the dimension grows and maintains its unlike nonstationary covariance function and therefore stationary covariance function is focused of this research. Exponential class, Matern class and rational quadratic are examples of stationary covariance functions. If the covariance is believed not to depend on the distance between the points in the input space but on the values the inputs take, one can consider non-stationary covariance functions. Example such as dot product and polynomial are the example of non-stationary covariance function and briefly described in [95].

The widely used squared exponential covariance function will be used here as the benchmark against which to compare alternative functions in terms of the model smoothness and fitting accuracy. Stationary covariance functions are the focus of this chapter in which power curve (described in chapter 4) is used as a key indicator since these are appropriate to wind turbine condition monitoring based on GP models. They are described as follows.

3.5.1 Exponential class covariance functions

Two types of covariance functions fall into the category of exponential class; i) the exponential covariance function (ECov) and ii) the squared exponential covariance function (SECov).

- **Exponential covariance function (ECov)**

The exponential covariance function (ECov) for any finite collection of inputs $\{x1, x2, \ldots\ldots, xn\}$ is defined as,

$$k_E(x, x') = \sigma_f \exp\left(-\frac{(x-x')}{l}\right) \tag{3.10}$$

Where, $(x - x')$ is the distance between x and x'(assuming $x > x'$), σ_f is the signal variance and l is the characteristic length. In order to include the effect of measurement errors associated with the SCADA data, an additional noise variance σ_n^2 is added to equation (3.10), which is then written as

$$k_E(x, x') = \sigma_f \exp\left(-\frac{(x-x')}{l}\right) + \sigma_n^2 \delta(x, x') \qquad (3.11)$$

Using equation 3.11, a GP model for a power curve is constructed using the GP theory described above, and shown in Figure 3.1 which indicates a clear case of overfitting and suggests that GP algorithm constructed using the exponential covariance function gives a fitting that tries to follow too much of the random variation in the data, and consequently is far from the desired smooth function that is expected for a wind turbine power curve. Note: the red line in the figure represents the confidence intervals, these are briefly discussed briefly in the upcoming section.

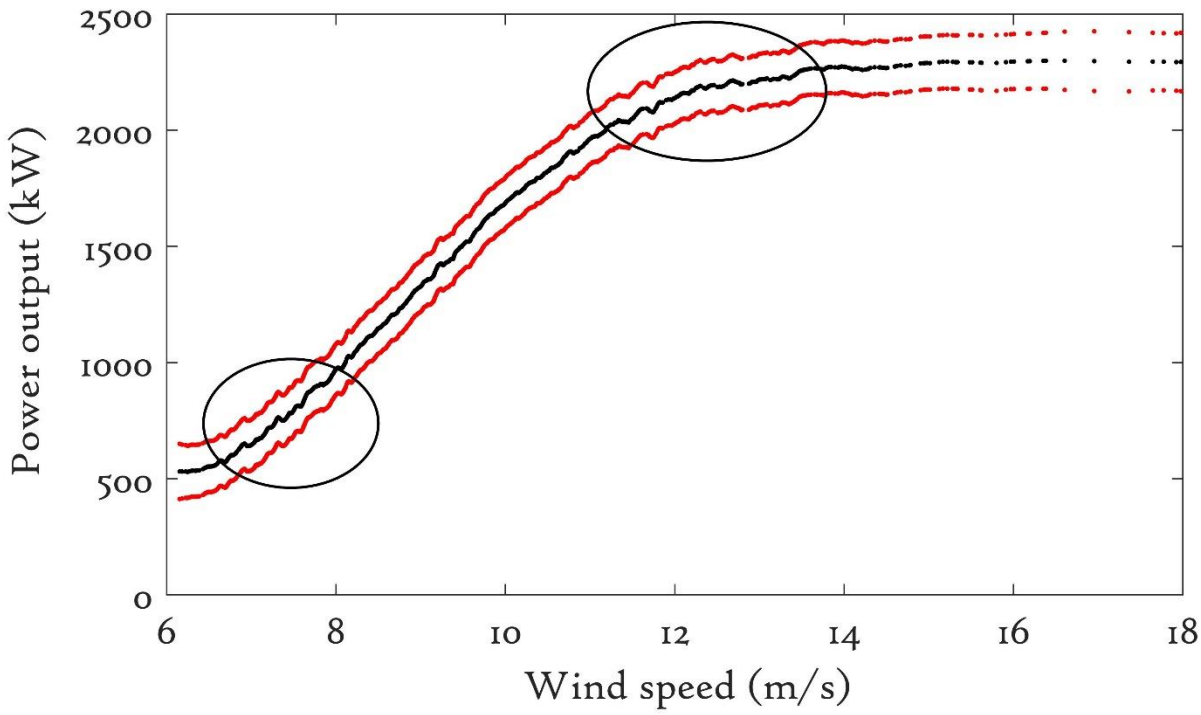


Figure 3.1: Power curve fitting using exponential covariance function

- **Squared exponential covariance function (SECov)**

The squared exponential covariance function (SECov) is a modified form of exponential covariance function and a widely used default covariance function in GP and SVM (support vector machine) models, [95,103]. For any finite collection of inputs $\{x1, x2, \ldots\ldots, xn\}$, it is defined as:

$$k_{SE}(x, x') = \sigma_f^2 exp\left(-\frac{(x-x')^2}{2l^2}\right) \quad (3.12)$$

where σ_f^2 , σ_n^2 and l are known as the hyper-parameters and suitable optimization of these parameters is the key to accurate GP model fitting.

Characteristic length scale (l) : Characteristic length scale (l) describes how quickly the covariance decreases with the distance between points. For example, a small value of length scale allows a rapid change in function values while larger values of length scale indicate constrain the rate of change in function values. In addition to these, the characteristic length scale (l) helps in determining how far we can reliably extrapolate from the training data.

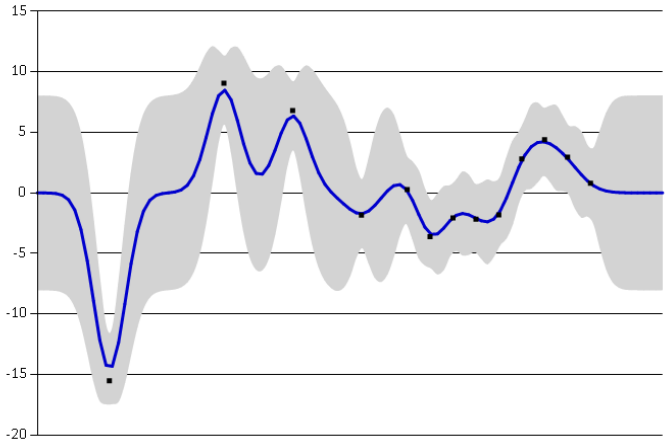


Figure 3.2: GP model with small length scale [103]

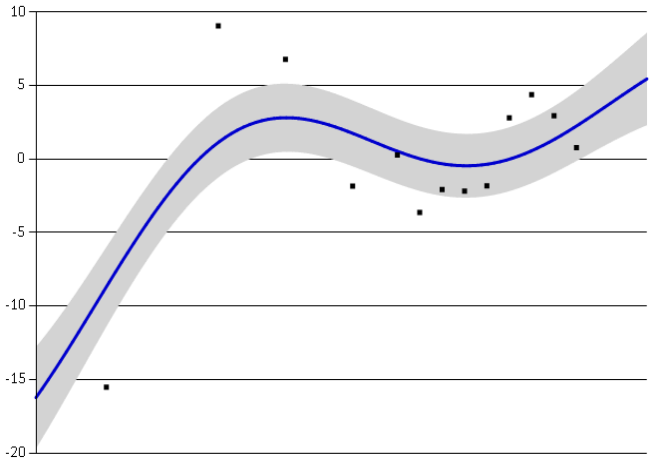


Figure 3.3: GP model with large length scale [103]

Signal variance (σ_f^2) : this is a scaling factor and signifies the extent of changes in functions values from their mean. For example, small values of σ_f^2 , characterize functions that remain close to their mean value while a large value of σ_f^2 allows more variation. If the signal variance is too large, then the modelled function will be free to encompass outliers, rather than highlight them as potential anomalies.

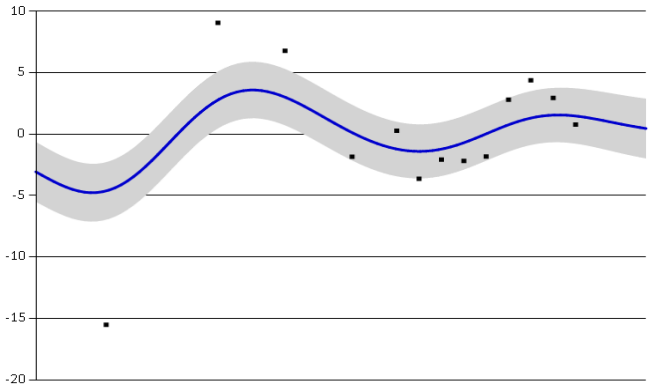


Figure 3.4: GP model with small-signal variance [103]

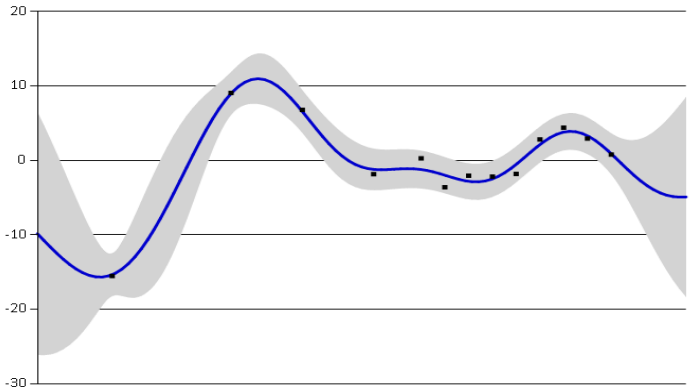


Figure 3.5: GP model with large signal variance [103]

Noise variance (σ_n^2) : Measurement data comes with noise. So in order to make GP model more accurate, noise variance added into the training data as mentioned above. The main function of noise variance is to specify how much noise is expected to be present in the data.

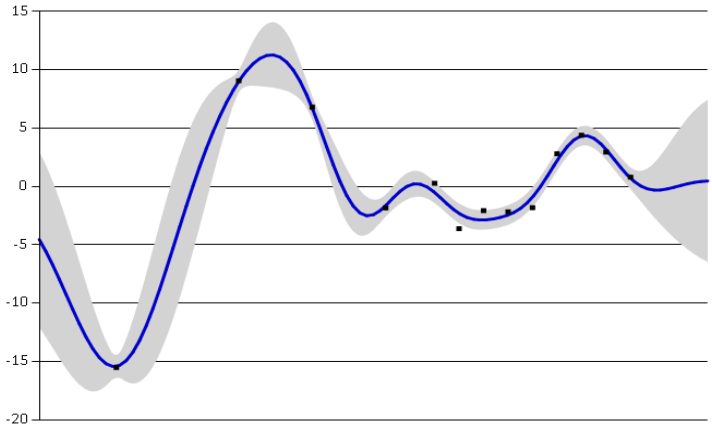


Figure 3.6: GP model with small noise variance [103]

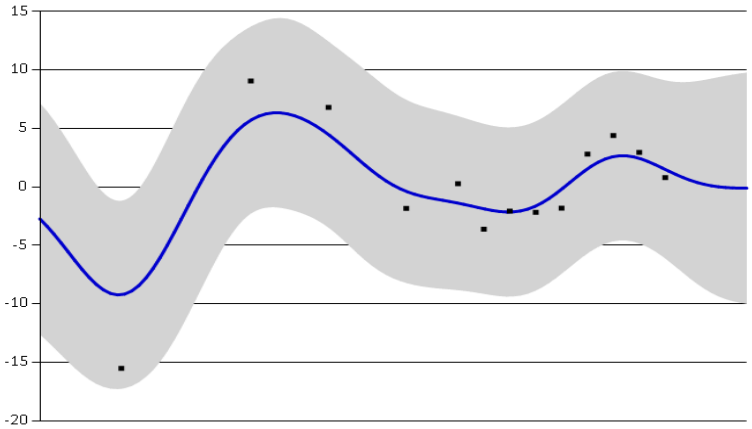


Figure 3.7: GP model with large noise variance [103]

The ref. [103] beautifully analysed the impact of length scale, signal variance, noise variance on a GP model and Figures 3.2 to 3.7 show why it is important to select suitable values of these parameters for GP model accuracy.

3.5.2 Matern class covariance functions

The Matern class of covariance functions was discovered by the Swedish statistician Bertil Matérn and is widely used in spatial statistics, [104], machine learning, [95], geostatistical analysis, [105], and other applications involving multivariate statistical analysis on metric spaces. It is commonly used to define the statistical covariance between observations made at two points that are r units distant from each other. The Matern function depends solely on r and hence is stationary in nature. The distance used here is Euclidean distance, hence the

Matérn covariance function is also spatially isotropic. Mathematically it is defined as,

$$k_v(r) = \sigma_f^2 \frac{2^{1-v}}{\Gamma(v)} \left(\sqrt{2v}\frac{r}{l}\right)^v C_v \left(\sqrt{2v}\frac{r}{l}\right) \tag{3.13}$$

where Γ is the gamma function, C_v is the modified Bessel function of the second kind, [106], and l and v are non-negative parameters.

If $v = l + \frac{1}{2}$, then the Marten covariance can be expressed as a product of an exponential and a polynomial of order l ,[107], as shown below,

$$k_{l+\frac{1}{2}}(r) = \sigma_f^2 exp\left(-\frac{\sqrt{2v}r}{l}\right) \frac{\Gamma(l+1)}{\Gamma(2l+1)} \frac{2^{1-v}}{\Gamma(v)} \sum_{i=0}^{l} \frac{(l+i)!}{i!(l-i)!} \left(\frac{\sqrt{8v}r}{l}\right)^{l-i} \tag{3.14}$$

If $v = \frac{3}{2}$ and $l = 1$ then equation (3.14), after simplification reduces to

$$k_{3/2}(r) = \sigma_f^2 \left(1 + \frac{\sqrt{3}r}{l}\right) exp\left(\frac{-\sqrt{3}r}{l}\right) \tag{3.15}$$

And if $v = \frac{5}{2}$ and $l = 2$ then equation (3.14), after simplification reduces to

$$k_{5/2}(r) = \sigma_f^2 \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2}\right) exp\left(\frac{-\sqrt{3}r}{l}\right) \tag{3.16}$$

where $r = \sqrt{(x - x')^T (x - x')}$ is the Euclidean distance between x and x'.

Equations (3.15) and (3.16) are the mathematical expressions for Matern 3/2 and Matern 5/2 covariance functions respectively.

Particular case: when $v \to \infty$, then the Matern function converges to the squared exponential covariance function, [107], as shown below.

$$\lim_{v \to \infty} k_v(r) = \sigma_f^2 \, exp\left(-\frac{r^2}{2l^2}\right) \tag{3.17}$$

Hence, the Matern covariance function includes the squared exponential covariance function as a particular case.

3.5.3 Rational quadratic (RQ) covariance function

The mathematical equation for rational quadratic covariance functions for points x and x' is defined below, [108],

$$k_{RQ}(x, x') = \sigma_f^2 \left(1 + \frac{(x-x')^2}{2\alpha l^2}\right)^{-\alpha} \quad (3.18)$$

Where α is a positive-valued scale-mixture parameter and determines the relative weighting of large-scale and small-scale variations. If we compare the rational quadratic function with the squared exponential covariance function, it is found that rational quadratic is equivalent to the adding a series of squared exponential covariance function with different values of the length scale, l. Thus, theoretically, it is expected that a rational quadratic covariance function fits the data smoothly across different length scales.

As already described in chapter 5, SCADA data comes with measurement errors so; it is desirable to add a noise term to the covariance function in order to improve the accuracy of the GP model. Hence equations (3.12), (3.15), (3.16) and (3.18) are modified to be:

$$k_{SE}(x, x') = \sigma_f^2 \exp\left(-\frac{(x-x')^2}{2l^2}\right) + \sigma_n^2 \delta(x, x') \quad (3.19)$$

$$k_{3/2}(r) = \sigma_f^2 \left(1 + \frac{\sqrt{3}r}{l}\right) exp\left(\frac{-\sqrt{3}r}{l}\right) + \sigma_n^2 \delta(x, x') \quad (3.20)$$

$$k_{5/2}(r) = \sigma_f^2 \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2}\right) exp\left(\frac{-\sqrt{3}r}{l}\right) + \sigma_n^2 \delta(x, x') \quad (3.21)$$

$$k_{RQ}(x, x') = \sigma_f^2 \left(1 + \frac{(x-x')^2}{2\alpha l^2}\right)^{-\alpha} + \sigma_n^2 \delta(x, x') \quad (3.22)$$

Here, σ_f^2 signifies the signal variance and l is the characteristic length scale. σ_n is the standard deviation of the noise term and impacts on model uncertainty, δ is the Kronecker delta, [95], and optimization of these hyper parameters was outlined in section 3.3.

3.6 GP power curve model based on covariance functions

Wind turbine SCADA datasets need to be filtered and air density corrected as per methodologies described in 2.5 and 5.4 to obtain accurate analysis. The SCADA data obtained after that are used to construct GP power curve based on different stationary covariance function described in section 3.3 together with the optimised parameters using MATLAB. The results are shown in Figure 3.8 together with an estimated 95% CIs and estimated power curve based different stationary covariance functions follow expected variance. This is further confirmed by plotting estimate power values against time series as shown in Figure 3.9. However, the real difference can be found in GP estimated uncertainty values which can be calculated and analyse by using CIs for individual GP power curve model. Finding a covariance function that reduces the uncertainty and increases early fault detection capabilities is the main objective and therefore significant comparative studies is being carried out in upcoming sections.

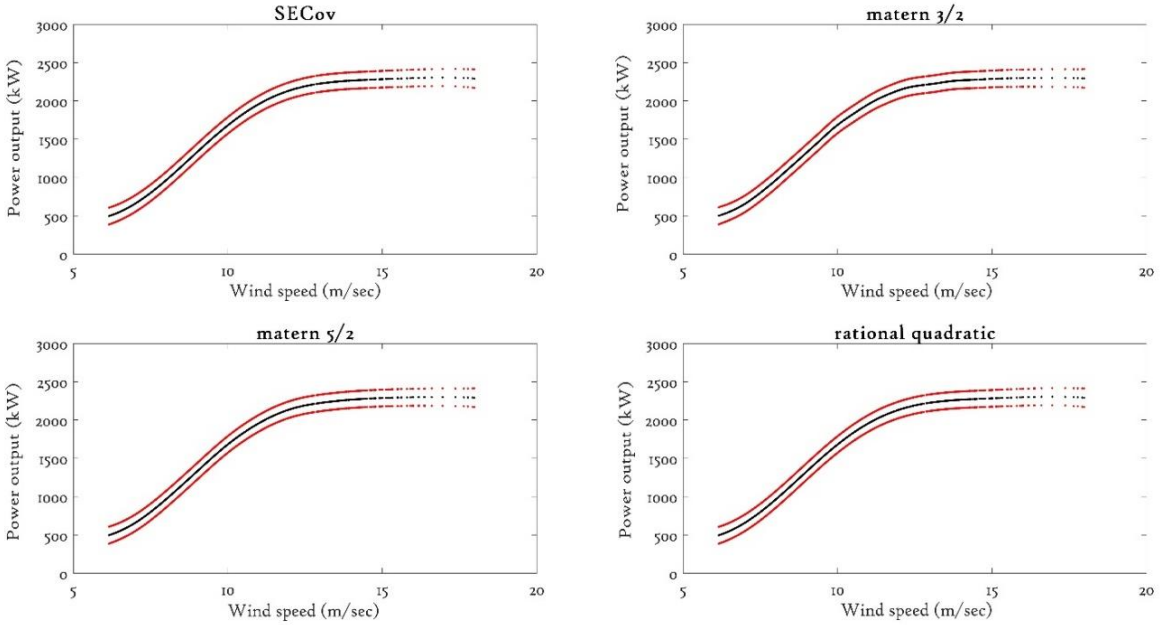


Figure 3.8: Modelled power curves with CIs for different covariance functions

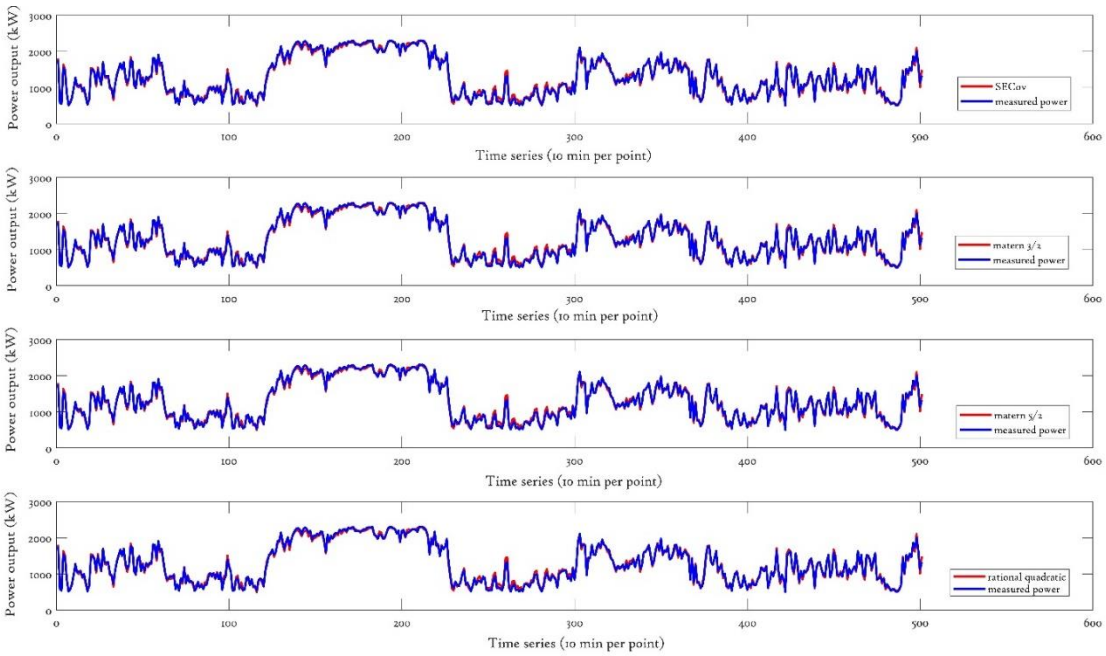


Figure 3.9: Estimated & measured power values

GP models uncertainty analysis using confidence intervals

In order to assess GP power curve accuracy, confidence intervals (CIs) can be used. CIs are a useful measure of uncertainty and the precision of model estimates and provide essential information about the uncertainty surrounding model estimates. For example, in chapter 5, a comparative analysis of uncertainty in IEC binned and GP fitted power curves using error bars and CIs respectively being carried out and found that uncertainty analysis of a power curve using CIs is relatively easy and straightforward as compared to binned error bars.

An accurate assessment of a GP power curve can be done using confidence intervals. These GP CIs provide information on the uncertainty surrounding an estimation but are itself a model-based estimate. Data points that lie outside of chosen confidence intervals can be considered anomalous, signifying a potential malfunction of the wind turbine. Also, uncertainty signifies the 'goodness' of the estimated fit. The CIs are a useful measure of uncertainty and thus the precision of model estimates. The standard deviation is the square root of the variance of

the predicted function ($Var[f*]$) (of equation 3.8) and used to estimate the CIs of the GP power curve model using equation (3.23),

$$CIs = \overline{f*} \pm 2\sqrt{Var[f*]} \tag{3.23}$$

Equation (3.23) suggest that CIs represent the pointwise mean plus and minus two times the standard deviation for given input value (corresponding to the 95% confidence region which represents the significance level of 0.05), for the prior and posterior respectively. CIs have been calculated using equation (3.23) and plotted as a function of the wind speed for covariance functions performance comparisons as described as follows.

As already mentioned, the GP model based on a squared exponential covariance function will be used as a benchmark in order to assess other GP models accuracy. Figures 3.10, 3.12 and 3.14 show the comparisons of fitted GP models based on different covariance functions for measured power curve data, while Figures 3.11,3.13 and 3.15 compare model uncertainty as a function of the wind speed quantified in terms of confidence intervals against the benchmark.

For the Matern 3/2 based GP model, the uncertainty is high, while with the Matern 5/2 based GP model it is small, but model fitting accuracy is less as compared to the benchmark GP model. The power curve fitting and uncertainty performance is almost the same for the benchmark model and GP model based on a rational quadratic covariance function, as shown in Figure 3.15.

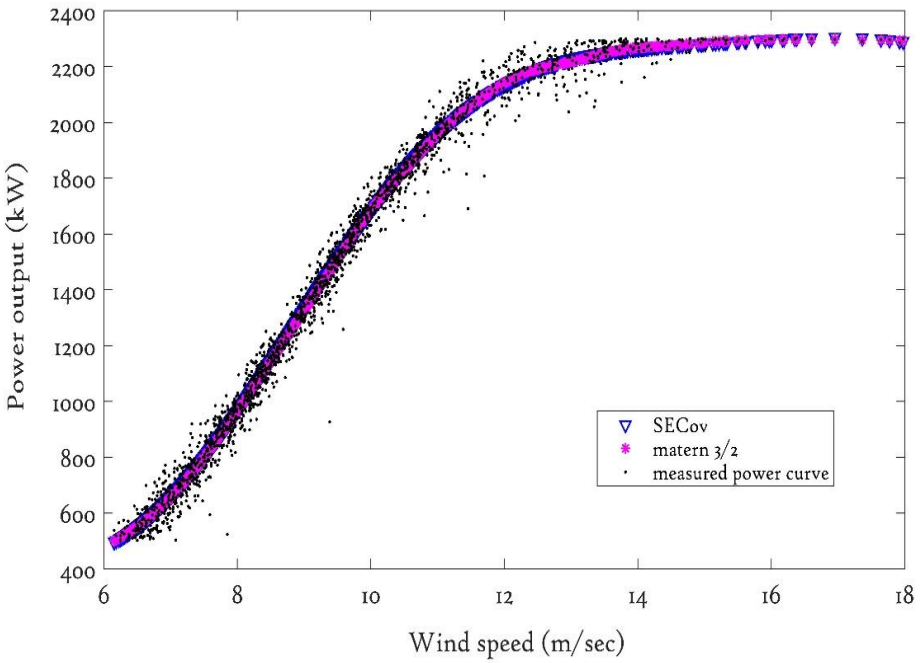


Figure 3.10: Estimated & measured power curve for SECov and Matern 3/2

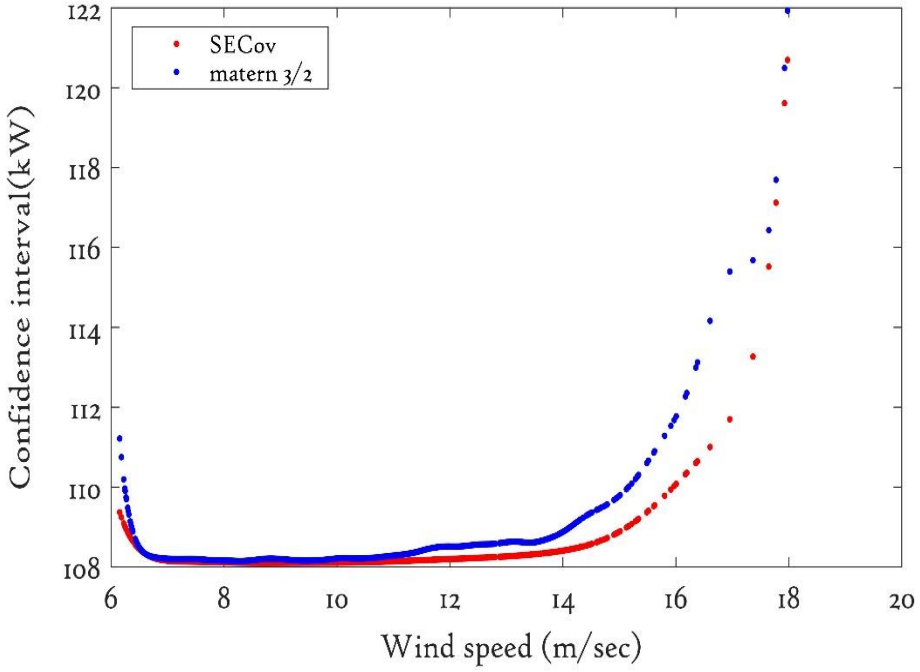


Figure 3.11: Uncertainty analysis using CI for SECov and Matern 3/2

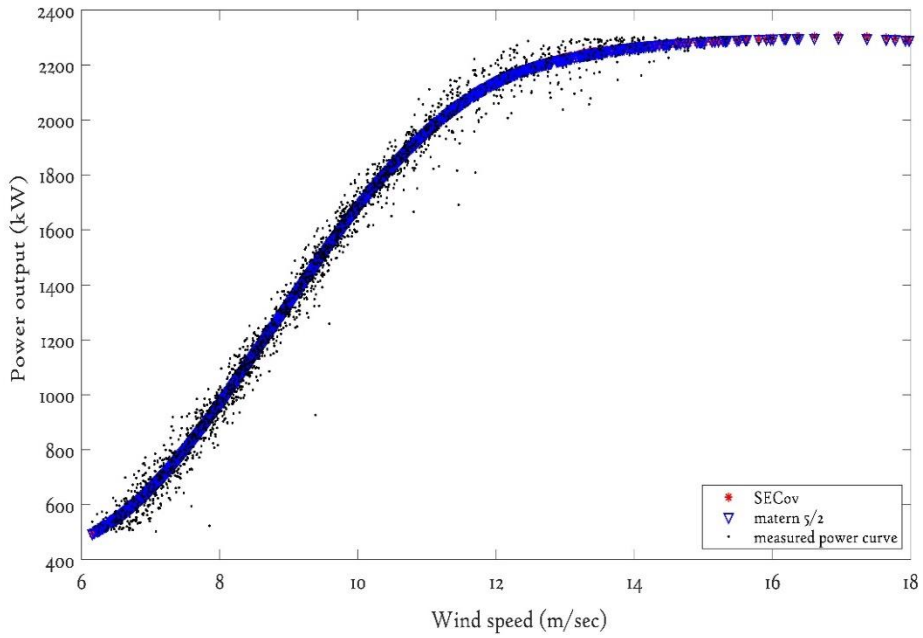


Figure 3.12: Estimated & measured power curve for SECov and Matern 5/2

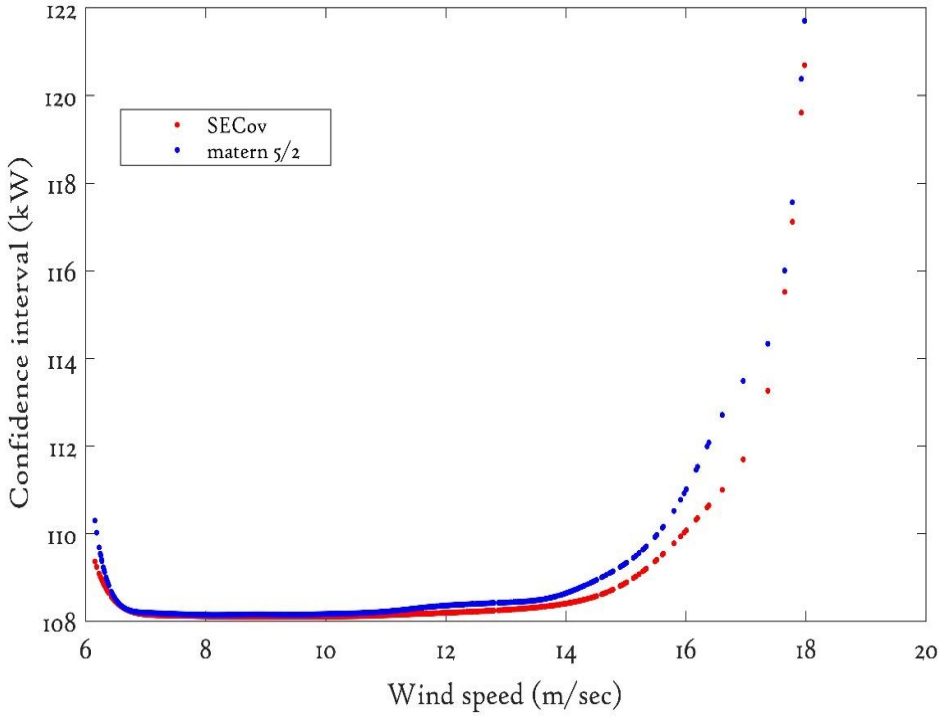


Figure 3.13: Uncertainty analysis using CI for SECov and Matern 5/2

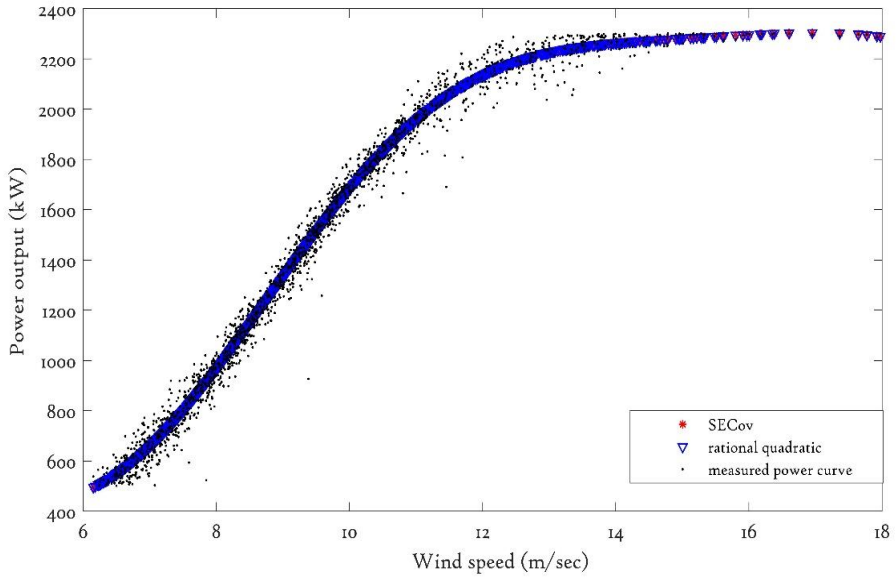


Figure 3.14: Estimated & measured power curve for SECov and rational quadratic

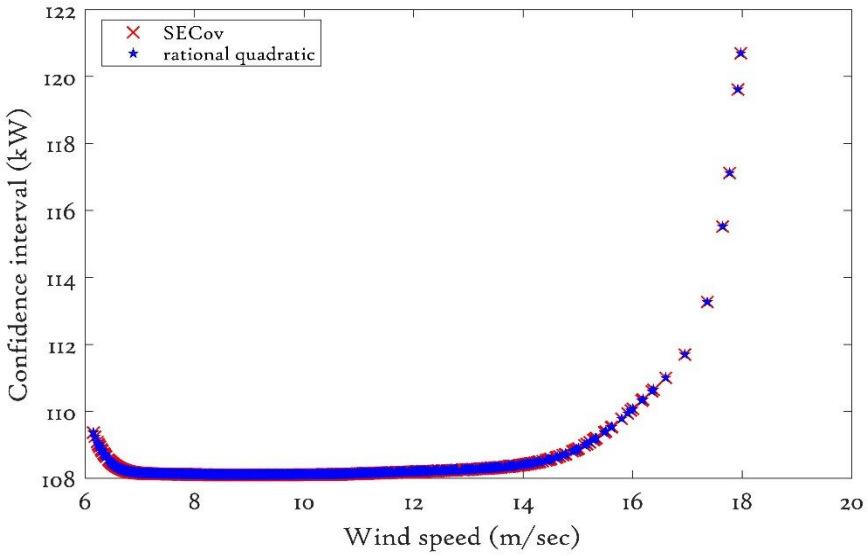


Figure 3.15: Uncertainty analysis using CI for SECov and rational quadratic

3.7 Performance comparisons

In this section, a comparative study of the different stationary covariance functions and their influence on GP power curve fitting is presented. Figure 3.16 shows the power curve of the different GP models with respect to the measured power curve and no significant differences between the individual GP models can be observed. Model uncertainty is compared through CIs, as shown in Figure

3.17. This suggests that the Matern 3/2 based GP model has relatively worst performance while both SECov and rational quadratic based GP models give the most accurate results. The GP based on Matern 5/2 performed better than Matern 3/2, but more impoverished than the SECov and rational quadratic GP models, especially in the region nearer to the rated wind speed range. Well above rated wind speed very little difference can be discerned between the models, as shown in Figure 3.17.

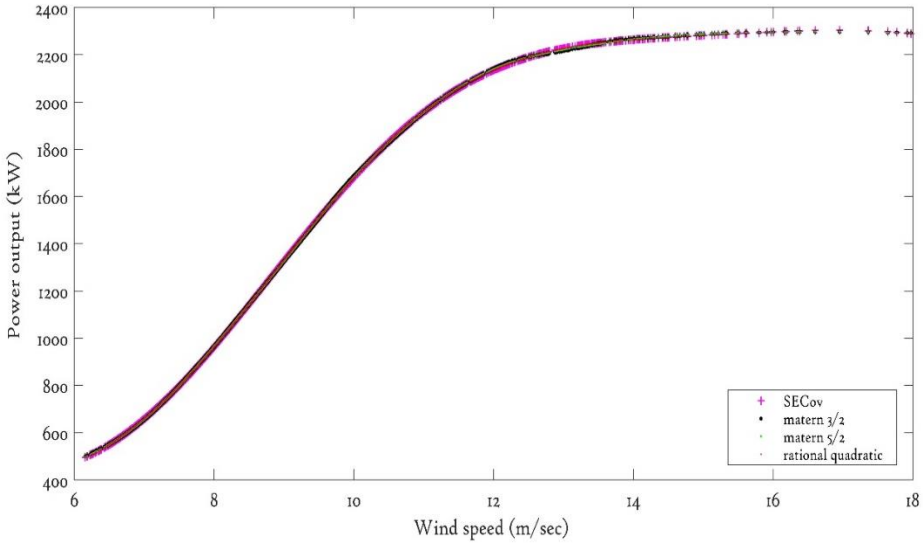


Figure 3.16: Comparative analysis of power curve fitting GP models

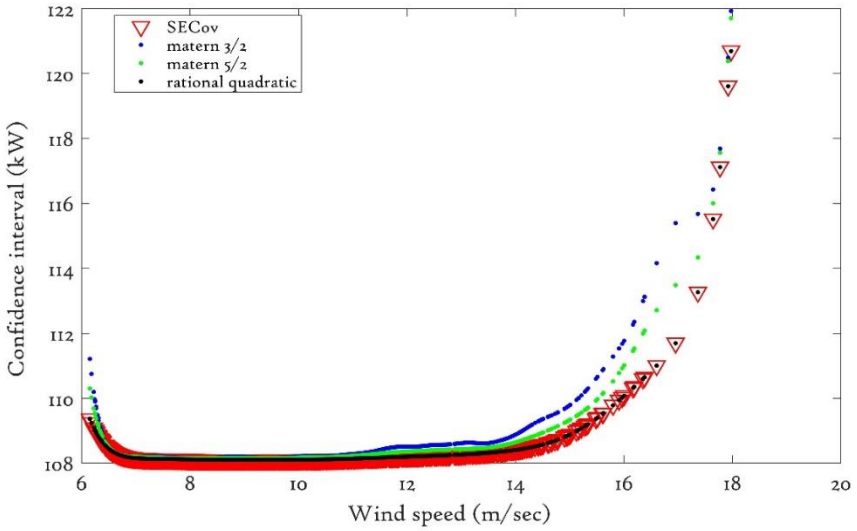


Figure 3.17: Stationary covariance function impact on GP model uncertainty via CIs

3.7.1 Using Residuals Analysis

Due to the nonlinear behaviour of a GP model, the residual which indicates the difference between measured and predicted values should also be investigated. Theoretically, residuals of a GP model should be Gaussian distributed, and this can be used to assess different stationary covariance function based GP power curve models for fitting accuracy and the nature of their distribution functions. The residual plots (Figure 3.18) indicates estimated GP values close to measured values and hence small model errors for all the GP models investigated. The frequency distribution of the residuals is shown in Figure 3.19 together with a fitted Gaussian distribution, and from that, it has been found that the rational quadratic and squared exponential covariance function based GP models have very similar distributions of residuals. Therefore, rational quadratic and SECov based GP gives similar histogram fitting and both are close to Gaussian (normal) distributions.

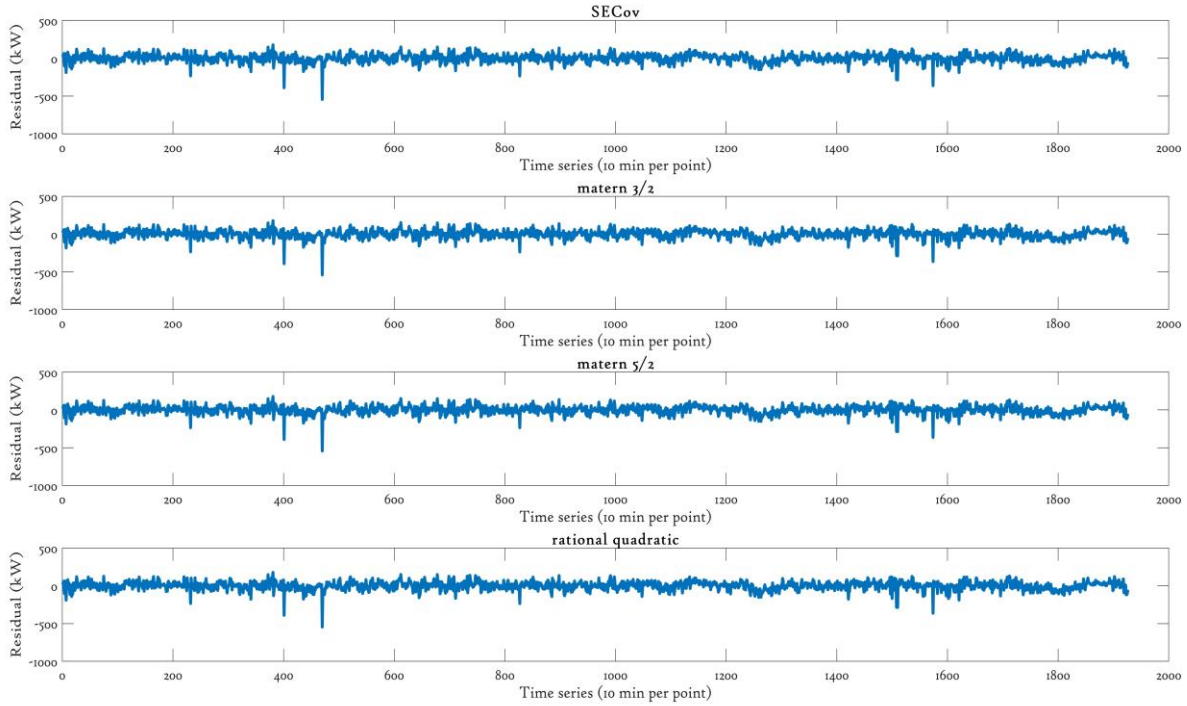


Figure 3.18: Residual plot for different GP model in time series

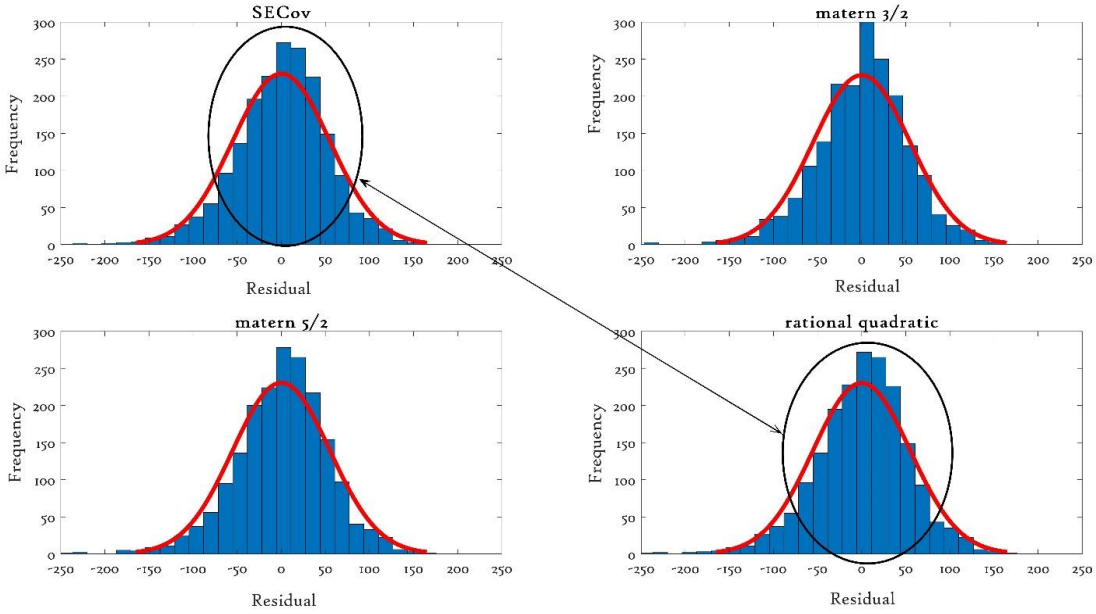


Figure 3.19: Histogram distribution fits for the different GP models

3.7.2 Using Performance Error Metrics

There are several statistical performance metrics that can be used to assess the performance of the estimated power curves such as the root-mean-squared error (RMSE), normalized mean absolute percentage error (NMAPE), symmetric mean absolute percentage error (sMAPE), the mean absolute error (MAE), and the coefficient of determination (R2), [109]. In this chapter, we use three goodness-of-fit indicators, namely the mean absolute error (MAE), the root mean squared error (RMSE), and the coefficient of determination (R^2) to evaluate the goodness-of-fit statistics of the GP power curve models. These are defined as follows. The difference between the measured and estimated values can be expressed by mean absolute error (MAE):

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} abs(y_i{}' - y_i)}{n} \tag{3.24}$$

, or in terms of residuals, $\mathrm{MAE} = \frac{\sum_{i=1}^{n}(e)}{n}$ (3.25)

To quantify the magnitude of the residuals (i.e., the difference between observed and modelled values, root mean square error (RMSE) is commonly used; defined as,

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i' - y_i)^2}{n}} \tag{3.26}$$

where y' are the GP predicted values for n different predictions, and y are the measured values. In terms of residuals this is:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(e)^2}{n}} \tag{3.27}$$

Another statistical measure, the coefficient of determination (R^2), [109], quantifies how close the data are to the fitted regression, and calculated as the square of the correlation between predicted output and measured values (hence always in the range from 0 to 1 with values closer to 1 indicates better fitting of the model to the data). It is defined as $R^2 = 1 - \frac{SSE}{TSS}$; where SSE is the sum of squared errors and TSS is the total sum of squares. Table 3.2 listed the computed figures of merit for the different covariance functions based GP power curve, and confirm that the performance of SECov and Rational quadratic is almost similar. However, prediction time and time to train the GP model is longer in case of rational quadratic covariance function as shown in Table 3.2.

MODELS	RMSE	R^2	MAE	PREDICTION SPEED	TRAINING TIME	REMARKS
GP SECov	**55.01**	**0.98**	**39.77**	**~26000 obs/sec**	**45.50 sec**	**Strong smoothness, faster**
GP Matern 3/2	54.83	0.99	40.46	~16000 obs/sec	59.33 sec	Poor smoothness
GP Matern 5/2	54.95	0.99	39.68	~17000 obs/sec	56.33 sec	Less smooth compared to SECov
GP Rational quadratic	**55.01**	**0.98**	**39.77**	**~20000 obs/sec**	**143.57 sec**	**Strong smoothness, slower**

Table 3.2: Figures of merit and computational time for the different covariance functions

3.8 Chapter conclusions

Gaussian Processes are a flexible class of models for which any finite-dimensional marginal distribution is Gaussian, and which can be viewed as a potentially infinite-dimensional generalisation of the Gaussian distribution. In this chapter, an in-depth analysis of stationary covariance functions has been undertaken. The power curve of a wind turbine is traditionally used as a critical indicator to assess the power performance of wind turbines, and hence comparative analysis of different covariance functions has been carried out are based on the power curve. The comparative analysis gives an idea which covariance function is suitable for GP fault detection algorithm and thus being discussed in this chapter. The results show that rational quadratic covariance functions based fitted power curve GP models perform almost same as the commonly used SECov functions although prediction speed and time is taken by model to train the model is slower, whereas GP model based on SECov performance is superior as shown in table 3.2. In short, both covariance functions work well in identifying a well fitted smooth function. The rational quadratic covariance function can be used instead of SECov for GP modelling if the data sets are small and there is no limitation on training time and prediction speed. While Matern 5/2 covariance can be used with caution, GP models based on Matern 3/2 perform poorly and are not recommended for wind turbine power curve modelling.

This chapter suggests that squared exponential covariance function is appropriate for wind turbine condition monitoring based on GP models and therefore in upcoming chapters squared exponential covariance function would be used for constructing robust GP models for wind turbine condition monitoring.

3.9 Chapter references

92. N. Wiener. Extrapolation, Interpolation and Smoothing of Stationary Time Series. MIT Press, 1949.
93. A. O'Hagan. Curve fitting and optimal design for prediction. Journal of the Royal Statistical Society B, 40:1–42, 1978C.
94. G. Matheron. The intrinsic random functions and their applications. Advances in Applied Probability, 5:439–468, 1973.
95. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X.
96. R. M. Neal. Bayesian learning for neural networks. In Lecture Notes in Statistics, 118, 1996.
97. Gaussian Processes A general overview plus discussion of the paper "Assessing Approximations for Gaussian Process Classification" by Malte Kuss and Carl Edward Rasmussen (from NIPS 2005). Presented by David P. Williams.
98. Joseph Gonzalez. Linear-Time Inverse Covariance Matrix Estimation in Gaussian Processes. 2017. doi: arXiv:1610.08035v4. (accessed 1st March 2018).
99. J. Hartikainen and S. Särkkä. Kalman filtering and smoothing solutions to temporal Gaussian process regression models. In: Proceedings of IEEE International Workshop on Machine Learning for Signal Processing, Kittila;2010, pp. 379-384. doi: 10.1109/MLSP.2010.5589113.
100. Simo Sarkka and Arno Solin et al. Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing. IEEE Signal Processing Magazine. 2013; 30 (4): 51–61. doi: 10.1109/MSP.2013.2246292.
101. Jie Chen and Nannan Cao. Parallel Gaussian Process Regression with Low-Rank Covariance Matrix Approximations. In: Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence (UAI2013); 2013,

Bellevue, Washington, USA. Available at https://arxiv.org/abs/1408.2060. (accessed 11th March 2018).

102. Ping Li and Songcan Chen. Gaussian Process Latent Variable Models. CAAI Transactions on Intelligence Technology.2016; 1(4): 366-376. doi: 10.1016/j.trit.2016.11.004.
103. Squared exponential kernel and its hyperparameters. Available online at http://evelinag.com/Ariadne/covarianceFunctions.html. (accessed 1st July 2018).
104. Noel Cressie, Hsin-Cheng Huang. (2012). Theory and Methods Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions. Journal of the American Statistical Association. 2012, 94(1999 - Issue 448), pp. 1330-1339.
105. Abhirup Datta, Sudipto Banerjee, Andrew O. Finley & Alan E. Gelfand. Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets. Journal of the American Statistical Association. 2016,111 (514): 800-812.
106. Frank Bowman. Introduction to Bessel Functions. (Dover: New York, 1958). ISBN 0-486-60462-4.
107. Abramowitz and Stegun. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. ISBN 0-486-61272-4.
108. David Duvenaud. The Kernel Cookbook: on Covariance functions. Available online at http://www.cs.toronto.edu/~duvenaud/cookbook/ (accessed 1st July 2018).
109. Glantz, Stanton A.; Slinker, B. K. Primer of Applied Regression and Analysis of Variance. McGraw-Hill. ISBN 0-07-023407-8,1990.

Chapter 4

Advanced nonparametric models performance comparison

The different types of statistical and nonparametric methods have been used to calculate the accurate power curves for wind turbine continuous performance assessment, forecasting and energy estimation; these are broadly classified into parametric and nonparametric methods in this chapter, three advanced nonparametric approaches, namely, Gaussian Process, Random Forest, and Support vector machine performance compared. The Gaussian Process, Support Vector Machine (SVM) and Random Forest (RF) are the advanced nonparametric methods that are used to modelled the wind turbine power curve using historical wind turbine SCADA data obtained from operational three bladed pitch regulated wind turbines.

The widely used statistical performance error metrics, uncertainty analysis, and residual analysis are taken into account to identify the most accurate power curve modelling approach. This chapter would highlight the strengths and weaknesses of the proposed advanced nonparametric techniques to construct a robust fault detection algorithm for wind turbines based on power curves.

4.1 Background and Motivation

Power production is the key consideration when assessing a potential site for wind farm development. The power output of a wind turbine is estimated from the power curve and wind speed profile for the site in question, and also from the site air density as this affects the wind power generation. The predicted long-term gross annual mean power output at a target site is calculated with the help of a Wind Turbine (WT) power curve. Moreover, power curve models can be useful in forecasting and capacity factor estimation purposes.

Numerous techniques have been introduced in the past to model WT power curves and these techniques generally divided into parametric and nonparametric. Parametric models are generally based on mathematical models that are often constructed from a family of functions with a number of variables that are fitted to correspond to the particular wind turbine. Widely used parametric approaches are segmented linear models [110], polynomial regression [111], and models based on probabilistic distributions such as four- or five parameter logistic distributions [112]. In contrast to the parametric approach, nonparametric approaches do not enforce any pre-specified condition, and thus, the estimated power curve is as close as possible to the measured data subject to the smoothness of the fit. Because of this, nonparametric models are able to model the power curve accurately over a wide range. Several studies have been conducted to develop an accurate power curve based on nonparametric techniques for performance evaluation, and these include a nonlinear regression model [113], cubic spline interpolation [114], neural networks [115,116], and data-driven methods (e.g., k-nearest neighbour clustering [117]). A comprehensive review of the existing WT power curve monitoring techniques can be found in [114].

GP models are used extensively in the literature [118,119] for a wide range of modelling applications; however, GP model applications are mostly for forecasting and prediction; in condition/performance monitoring, their application has to date been limited. SVM is another nonparametric method that has been introduced for wind turbine power curve modelling [120,121]. However, both methods suffer from a number of practical issues such as the cubic inversion issue associated with massive datasets (see chapter 3). Finally, the RF model is another nonparametric approach used to construct power curves. Unlike most of the nonparametric approaches, the RF does not need to be tuned or optimised, and it incorporates the prediction of several weak predictors [122]. As the name suggests, it is used to create a forest in some way and make it random while maintaining the direct relationship among the number of trees in the forest. Usually, a large number of trees indicates a more accurate result. It is worth noting that RF and decision tree techniques are not the same because with the RF random samples are used to obtain the root node and splitting the feature nodes, in contrast to decision trees [123]. These advanced nonparametric models are flexible and easy to implement, and computationally straightforward to implement; thus comparative studies of GP, SVM and RF needed to find the best approach among them for wind turbine condition monitoring.

4.2 Chapter novel contributions

As has already been mentioned, power curve model accuracy varies with the techniques and the particular data set used, and hence no single technique performs best in all cases of observations obtained from different WTs. It is therefore essential to investigate the performance of different nonparametric techniques for power curve modelling to evaluate which technique is more accurate for a given dataset. Advanced nonparametric models such as the GP, SVM, and RF are gaining popularity because of their low computational cost and high accuracy. The direct comparison between these models can be useful in

identifying the method that is more robust and computationally feasible. This chapter aims to fill this gap.

The chapter presents the implementation of three advanced nonparametric algorithms (GP, SVM, and RF) for modelling of wind turbine power curves, and their accuracy has been compared using error performance metrics. Performance comparison of these methods is made to identify the best approach taking into account the computational cost and required processing power. The SCADA dataset obtained from modern pitch regulated wind turbines is used to train and validate the performance of the proposed nonparametric models. The outcomes should be useful in constructing power curve based fault detection algorithms, where accuracy and uncertainty are paramount, for wind turbine condition monitoring. A framework for modelling wind turbine power curves and its performance comparisons is presented in Figure 4.1 and described as follows. The SCADA data obtained from operational wind turbines are collected which is then filtered and air density corrected (described in chapter 5). After this, datasets divided into training and validation; training data points are used to train the models, and validation data points are used to validate the performance of models. Performance Error metrics, residuals analysis, and uncertainty analysis are used to compare the performance of the models and based on this comparison, the best approach for wind turbine power curve modelling is being suggested.

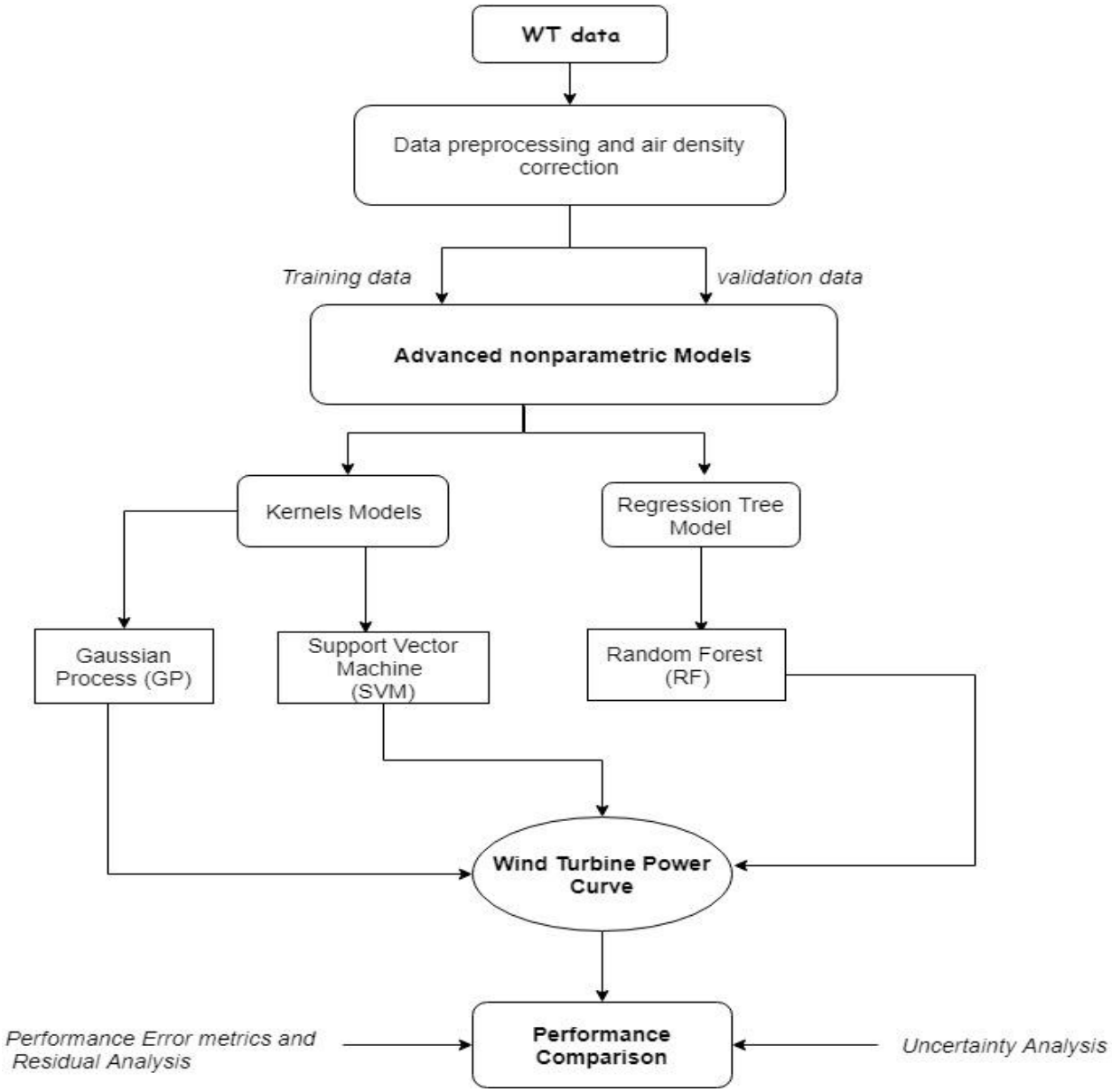


Figure 4.1: A framework of the advanced nonparametric Wind Turbine Power Curve models for performance comparison

4.3 Why wind turbine power curve?

The wind turbine power curve described the nonlinear relationship between the power output and hub height wind speed, see Figure 4.2. The electrical power output of the turbine is not only dependent on wind speed but also affected by turbulence intensity, wind direction, vertical and horizontal shear, drive train temperature, yaw error and so on [124].

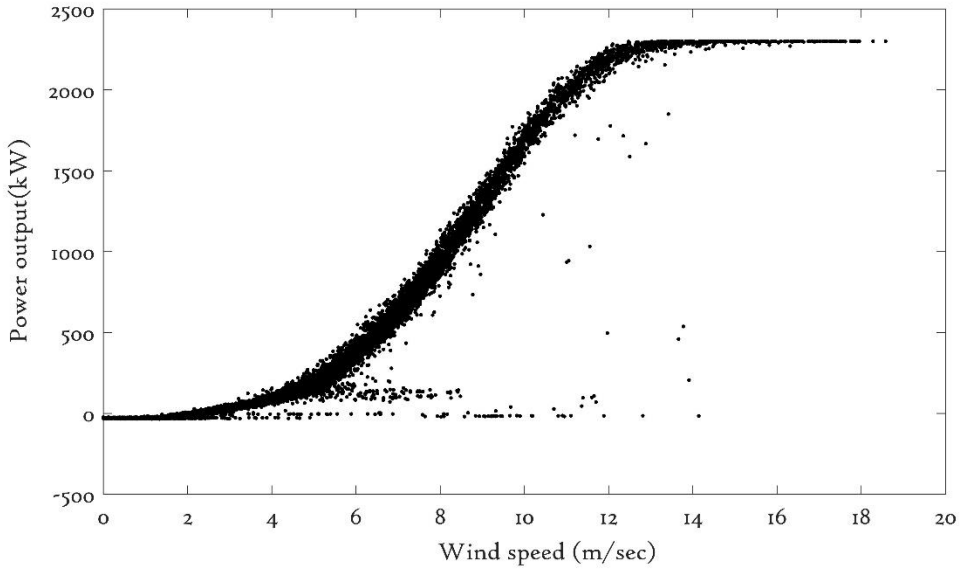


Figure 4.2: Measured power curve of an industrial wind turbine

An accurate power curve is not only used to improve performance assessments but can play a significant role in identifying different wind turbine fault types [74]. Usually, an individual WT has unique power curve depending to the operating conditions for which it has been designed (e.g., wind speed range) and actively used for continuous monitoring the performance by differentiating between a normal and an abnormal state [124]. The WT power curve follows the sigmoid shape and any changes in its characteristic shape likely to indicate abnormal operation due to a fault. The wind speed between the cut-in and the cut-out speed ranges are considered significant because this operational region presents a significant opportunity to optimise the power generation process, described briefly further in upcoming chapter 5.

The SCADA datasets of turbine rated at 2.3 MW of wind farm located in Scotland, UK has been used in this research to evaluate the performance of advanced nonparametric models. The total data points of 13,250 that begin with time stamp ‘‘1/10/2012 00:00 AM’’ and ending at time stamp ‘‘31/12/2012 23:50 PM’’ are sampled at 10-minute average and are used for models training and validations. These measured data points became 3960 data points after pre-processing (Table 4.1) as per the methodology described in section 2.5. In

addition to that, IEC Standard air density approach outlined in 5.4 are used to make power curve air density corrected. Figure 4.3 signifies the air density corrected and pre-proceed measured power curve. The whole SCADA data are divided into two parts; training data (70% of all of the data) and validation data (30% of all of the data). Training SCADA data used to train the model while validation SCADA data are used to validate the models' accuracy.

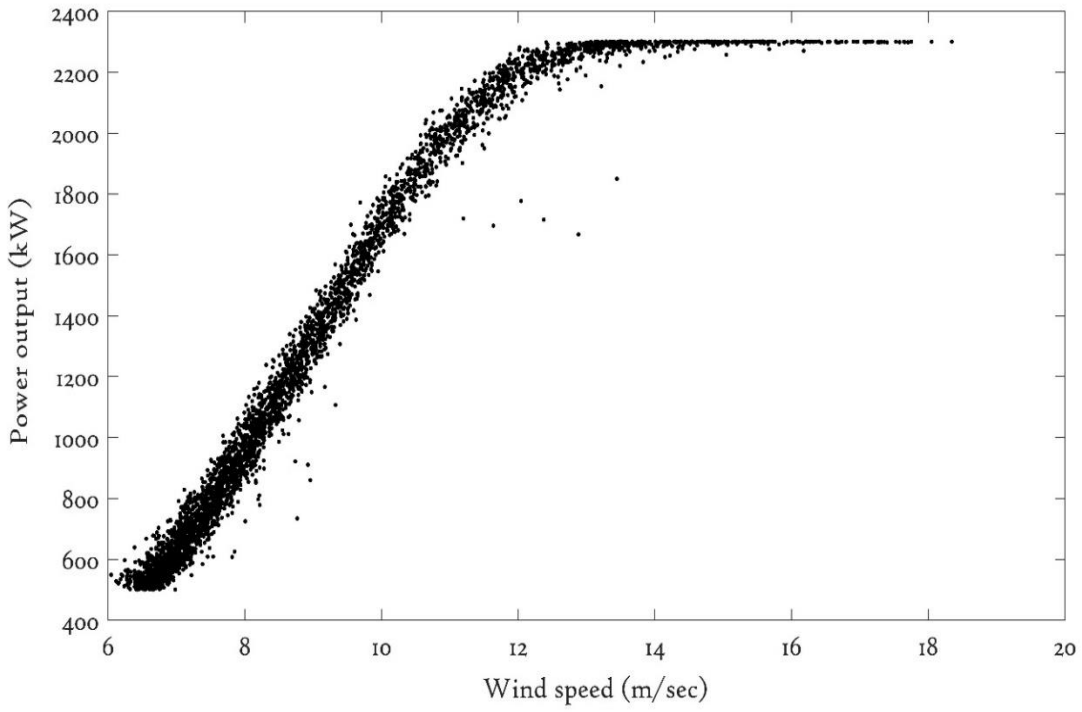


Figure 4.3: pre-processed and air density corrected power curve

Start timestamp	End timestamp	Measured data	Filtered data	Training data	Validation data
1/10/2012 00:00 AM	31/12/2012 23:50 PM	13250	3960	2500	1460

Table 4.1: SCADA data description

4.4 Power curve modelling using advanced nonparametric models

In this section, the algorithmic procedures of the proposed advanced nonparametric power curve modelling approaches are explained in detail. The three advanced nonparametric models, namely; GP, SVM, and RF were used to construct the power curve of a WT. Out of these three approaches, GP and SVM are kernel based methods while RF technique is based on a regression tree.

4.4.1 Power curve model based on Gaussian Process

GP models theoretical description to estimate the values are well described in chapter 3 and are used to power curve. The training SCADA data of Table 4.1 is used to train the GP model while validation SCADA datasets are used to test the accuracy of the GP model (in MATLAB) and result shown in Figure 4.4. Figure 4.4 suggests that GP power curve is smooth and continuous and, able to fit the measured power curve accurately.

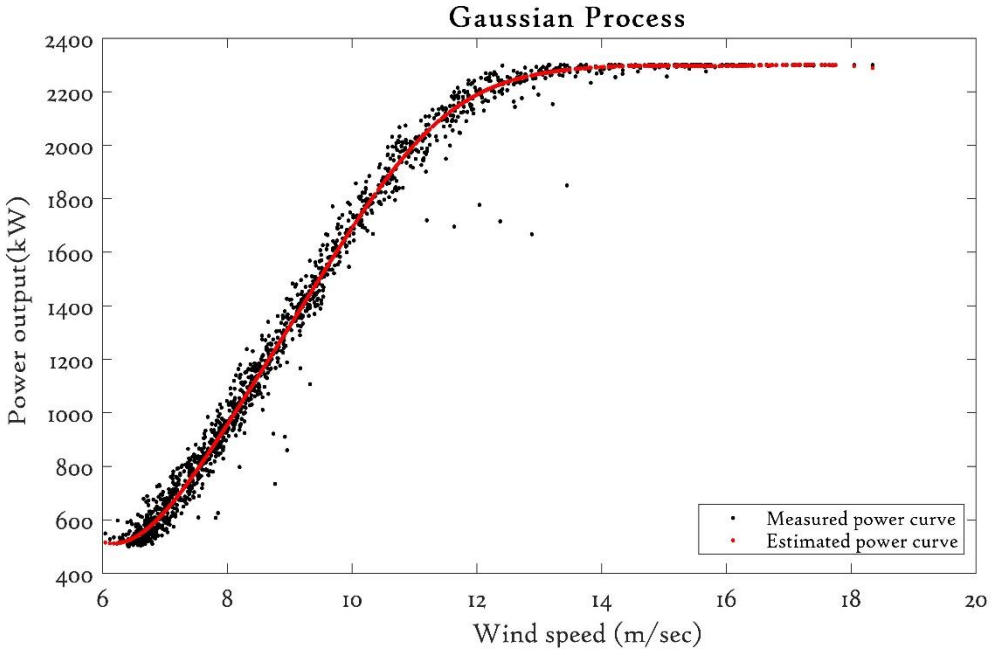


Figure 4.4: Gaussian Process based power curve

4.4.2 Power curve model based on Support Vector Machine

The SVM is a nonparametric, machine learning technique, which follows the principle of structural risk minimisation. SVM mostly used in solving a problem related to classification (called support vector classification) and regression (called support vector regression). SVM based on statistical learning theory and Ref. [125,126] provide a detailed explanation of it.

In this study, epsilon-insensitive SVM (ε-SVM) regression used (i.e., *L1* loss) where the dataset includes predictor variables and observed response values. In ε-SVM regression, the primary objective is to find a function $f(x)$ that differs from y_n to an extent not greater than ε for each training point x and be flat at the same time as much as possible. To model power curve, the dual formula for nonlinear SVR used since dual formulation lends itself easily to classify data which is not linearly separable in the original feature due to the 'Kernel Trick' which makes computations implicitly in high dimensional space. In 'Kernel trick' the inner product in the input space is replaced by kernel function and this approach is particularly valuable in a condition where it is more convenient to compute the kernel than the feature vector themselves.

The dot product $x_1'.x_2$ (of linear SVR: dual Formula) replaced by a nonlinear kernel function, $G(x_1, x_2) =< \varphi(x_1), \varphi(x_2) >$ where $\varphi(x)$ is a transformation that maps x to a high- dimensional space and is called dual formula and being used in this research. The Gram matrix is an n-by-n matrix which contains elements; $g_{ij} = G(x_i, x_j)$, where each element $g_{i,j}$ is equal to the inner product of the predictors as transformed by φ. It should be noted that using kernel function, Gram matrix can be calculated directly hence don't need to know the φ value. Using this technique, nonlinear SVM calculates the optimal function f(x) in the transformed predictor space. Using nonnegative multipliers (α_n and α_n^*), the Lagrangian function of the primal function constructed for each observation x_n

which leads to dual formula, where a nonlinear regression finds the coefficients that minimise,

$$L(\alpha) = 0.5 \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\, G(x_i, x_j) + \varepsilon \sum_{i=1}^{N} (\alpha_i + \alpha_i^*) - \sum_{i=1}^{N} y_i (\alpha_i - \alpha_i^*) \quad (4.1)$$

Under the following constraints:

$$\sum_{n=1}^{N} (\alpha_n - \alpha_n^*) = 0 \; ;$$

$$\forall n : 0 \leq \alpha_n \leq C \; ;$$

$$\forall n : 0 \leq \alpha_n^* \leq C.$$

In SVM regression, the input (x_n, x) first mapped onto a m – dimensional feature using fixed nonlinear mapping, and then a linear model is constructed in this feature space using a following mathematical equation,

$$f(x) = \sum_{n=1}^{N} (\alpha_n - \alpha_n^*)\, G(x_n, x) + b \quad (4.2)$$

where $f(x)$ is the linear model (in the feature space) and this particular technique called ε-SVR regression [127]. The ε -insensitive loss function is used to build the objective function of the ε- SVR.

The Karush-Kuhn-Tucker (KKT) complementarity conditions are optimisation constraints required to obtain optimal solutions. For nonlinear SVM regression,

$$\forall n : \; \alpha_n \left(\varepsilon + \xi_n - y_n + f(x_n)\right) = 0 \; ;$$

$$\forall n : \; \alpha_n \left(\varepsilon + \xi_n^* + y_n - f(x_n)\right) = 0 \; ;$$

$$\forall n : \; \xi_n (C - \alpha_n) = 0 \; ;$$

$$\forall n : \; \xi_n^* (C - \alpha_n^*) = 0 \; ;$$

These conditions indicate that all observations strictly inside the epsilon tube have Lagrange multipliers $\alpha_n = 0$ and $\alpha_n^* = 0$. If either α_n or α_n^* is nonzero value zero, then the corresponding observation referred to as support vectors.

The C is the box constraint which control the penality imposed on the observation that lie outside epsilon margin (ϵ) and help to prevent model overfitting. This value determines the trade-off between the flatness of function $f(x)$ and the degree to which larger deviations (larger than ε) are tolerated. In short, both *C* and ε values affect SVR accuracy, and therefore it is necessary to find optimal values for these parameters by using appropriate optimization techniques. The optimization of SVM regression based on Sequential minimal optimization (SMO) algorithm and Cross validation is being carried out to calculate the best possible value for box constraint and epsilon based on the work of [126,128].

The calculation of ε and C are based on the nature of input datasets and the choice of kernel. In this study, a Gaussian kernel was used and *C* values is calculated as iqr(Y)/13.49 where iqr(Y) is the interquartile range of the response variable Y, [129]. The 13.349 is a rescaling factor (that quantifies the statistical dispersion in a set of numerical data) that reflects the change from interquartile range to standard deviation. The bias $\in$ is a part of the original primal formula (equation 4.1) of the SVR and is calculated from the following equation, [126,128],

$$\in = \sum_{n=1}^{N}(\alpha_n - \alpha_n^*) \tag{4.3}$$

Where α_n and α_n^* are the nonnegative multipliers for each observation x_n. The obtained biased value added into the model to predict the power curve of wind turbine accurately.

The Gaussian kernel is also popularly known by radial basis function (RBF) kernel because it makes computation faster and involves computations in higher dimensional space. In this chapter, the Gaussian kernel is used model SVR based power curve and mathematically expressed as,

$$k(x, x') = exp(-\gamma\|x - x'\|^2) \tag{4.4}$$

Where γ is the kernel scale for given points x and x'.

To select the best kernel scale parameter, cross-validated 'grid-search' cross-validation carried out for RBF kernel of the SVM regression for power curve modelling in which 10-fold used.

The SCADA dataset of Table 4.1 was used to estimated the power curve and result is shown in Figure 4.5. The SVR based power curve is continuous and accurately predicts the measured power curve. However, at above-rated wind speed, the SVR power curve accuracy deteriorates because of lack of a sufficient number of data points in that wind speed region. Furthermore, it suffers from the cubic inversion issue like GP models (described in chapter 3) which affect the model accuracy and caused heavy computational load while dealing with large datasets.

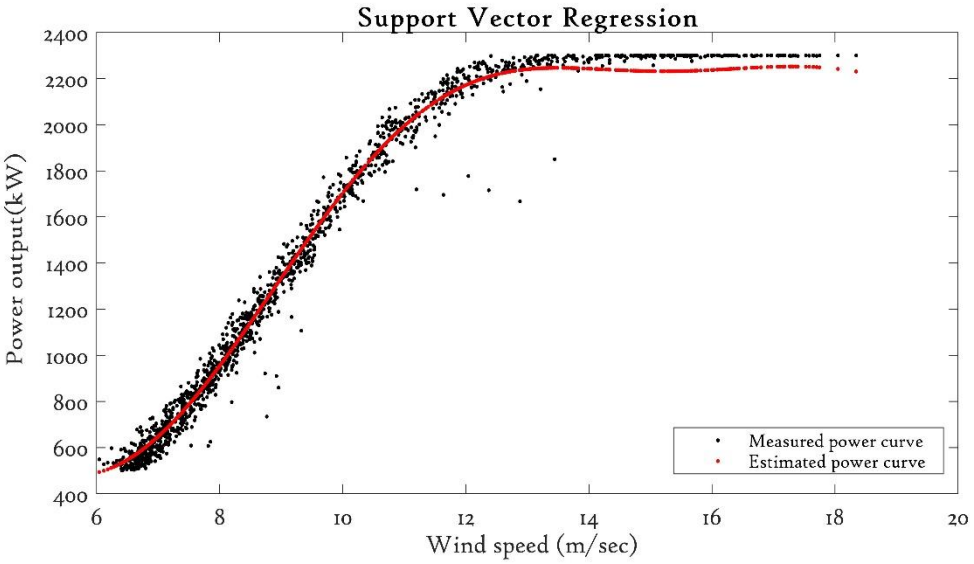


Figure 4.5: Support Vector Regression based power curve

4.4.3 Power curve model based on Random Forest

Random Forest (RF) is a nonparametric, supervised learning approach that combines the prediction of several decision trees to get a more accurate and stable estimation. The RF creates a forest which is an ensemble of decision trees (hence RF called Ensemble learning technique) and makes it somehow random [123] and these trees are mostly trained with the 'bagging' method. The systematic and

detailed explanation of RF can be found in [129]. Here, a brief description of RF would be provided. The RF is a collection of Classification and Regression Trees (CARTs) in which CART splits the input space recursively, according to a predefined split criterion, to small rectangular regions and then fits a simple model, commonly a constant value, in each one of them, and this can be demonstrated by the tree diagram, see Figure 4.6.

The fundamental principle is called bagging (bootstrap aggregation); where a sample of size n taken from the training set S_n is selected randomly and fitted to a regression tree and grow the tree using CART approach to maximum size and do not prune.

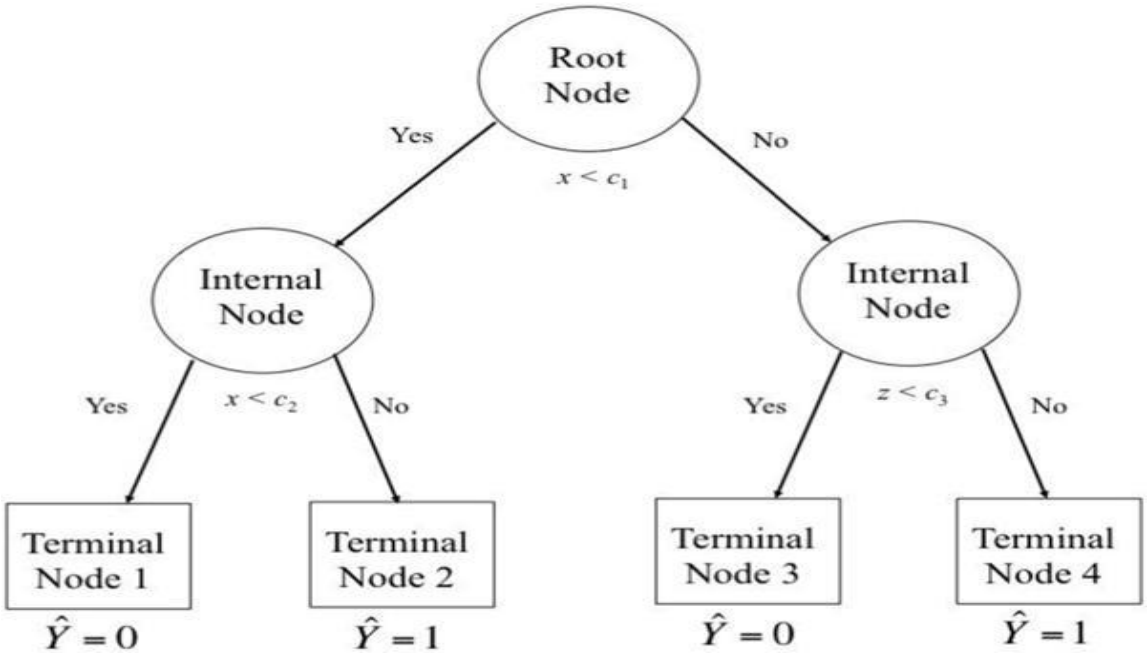


Figure 4.6: Example of tree diagram from classification and regression tree (CART) analysis, [130]

In RF ensemble learning approach, a group of 'weak learners' used together to form a 'strong learner' to improve the performance. RF uses decision tree in which each tree is constructed from a bootstrap sample from the original dataset with an objective to increase diversity between members of the ensemble by restricting classifiers to work on different random subsets of the full feature space [131]. In the RF approach, k bootstrap sampled randomly and then a regression

tree fit on each sample. After that, the average values of k regression tree are taken in order to make an estimation.

In this section, RF algorithm as per [132], used to estimate the power curve of a WT where bootstrap samples are generated similar to bagging algorithm. The bootstrap aggregation produces non-correlated trees through different training samples which give immunity to noise. However, instead of using all training data to fit the tree, only random predictor variables are used at each split. Splitting the decision improve the RF accuracy such that the reduction in the residual sum of squares is maximised [132]. Here, SCADA datasets (of Table 4.1) divided at the first node where all variables (wind speed and power output) values are considered, and after the split, further variable and splitting condition is selected and this repeated again. While doing this, the same variable can be selected consecutively and hence this splitting technique called recursive binary splitting. To find the optimal values, randomly selected predictor (k_{tr}) can vary.

It should be noted that RF tried to search for the best split among the k_{tr} selected features and this selection is uniform. The randomly selected predictor (k_{tr}) is same for all prediction trees and it is recommended to be the square root or one third of the features number k as: $k_{tr} = \sqrt{k}$ or, $k_{tr} = \frac{k}{3}$. After that, the RF algorithm is similar to the CART where by minimizing the cost function, best split is obtained and repeat the procedure until full development of all trees. RF models are very good at capturing the nonlinear relationship between features and the target and in minimizing the overfitting issue. The spatial 10-minute average training SCADA datasets (Table 4.1) are used to estimate the power curve based on RF and is shown in Figure 4.7. Figure 4.7 suggest that the RF based power curve is accurately following measured power curve variance but neither it is continuous or smooth. It is worth to note that, the RF power curve is a predictive model, not a descriptive model and hence it does not give a description of the relationship among the predictors. Moreover, confusing data makes RF

inaccurate and confusing, hence it is desirable to select appropriate and error free predictors that affect the target variables.

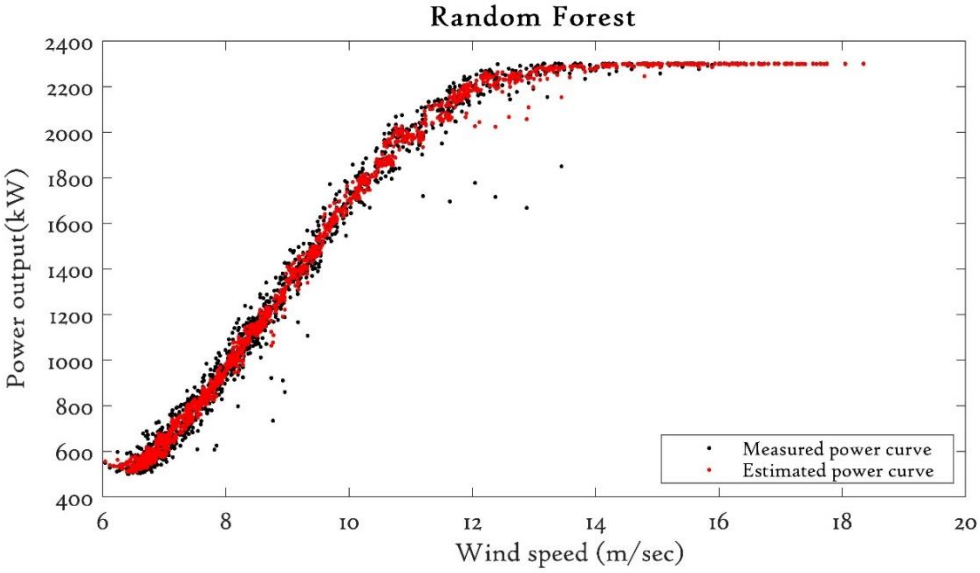


Figure 4.7: Random Forest based power curve

4.5 Performance comparisons of the modelled power curve

The advanced nonparametric models discussed in Section 4.4 are used for comparative analysis in order to find out which of the proposed advanced nonparametric models is able to reflect the dynamic properties of power curves accurately and what are advantageous and disadvantageous of proposed advanced nonparametric models. To do this, residual analysis, performance error metrics are used for performance comparison. The advanced nonparametric models result presented in above are compared in Figure 4.8 together with the measured power curve and comparative analysis suggest that GP based power curve is relatively more accurate and has a continuous and smooth fitting which closely following the expected variance at all wind speed range while RF based power curve is neither continuous nor smooth because it is built on CART theory, but closely matching the measured power curve. The performance of the power curve based on SVM deteriorates after rated wind speed because of the unavailability of reasonable numbers of SCADA data points as shown in Figure 4.8.

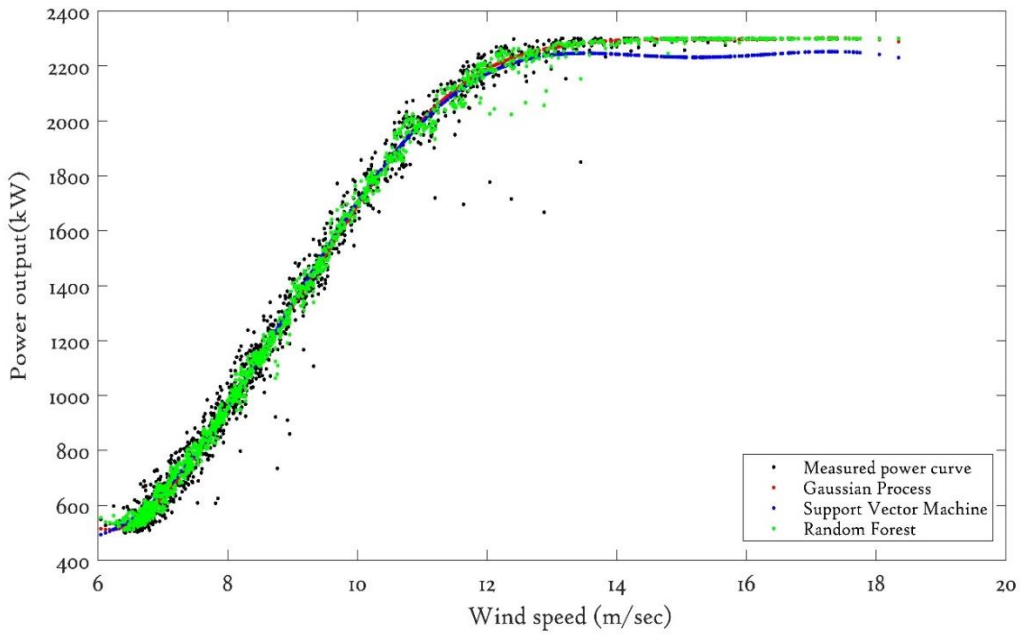


Figure 4.8: Comparative analysis of nonparametric models

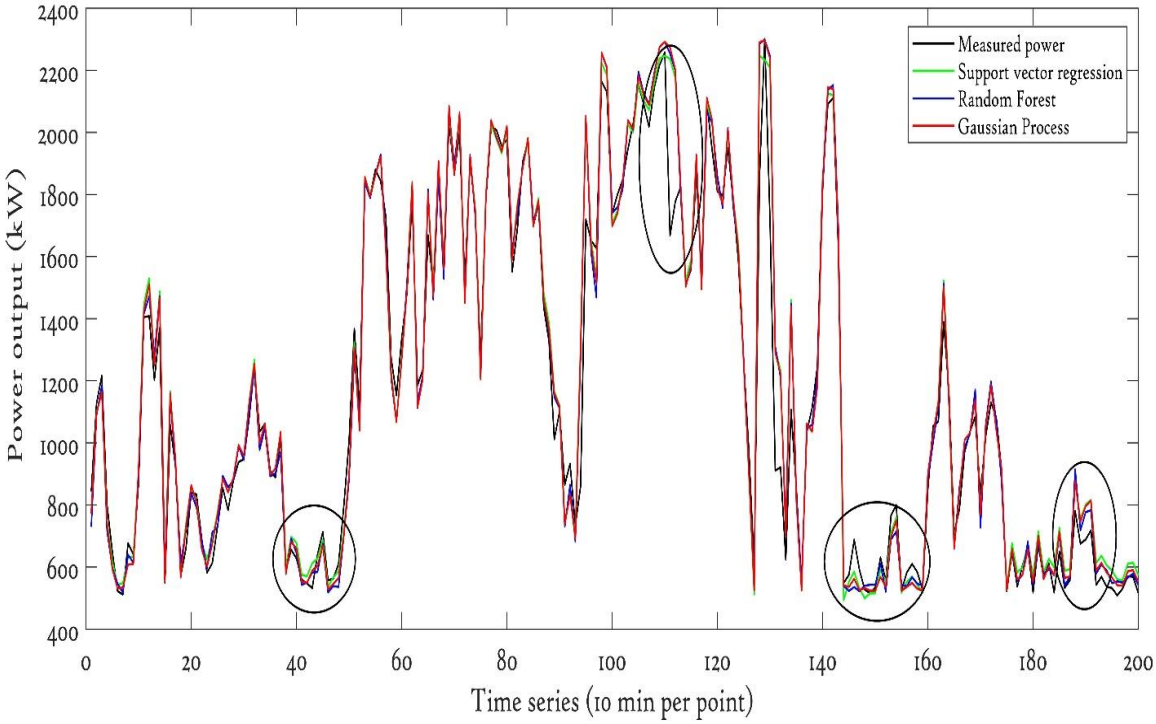


Figure 4.9: Comparative analysis of nonparametric models in terms of time series

Figure 4.9 shows the estimated power values in the time series of proposed advanced nonparametric models and signifies that the GP model able to estimate power most accurately in the entire time series.

4.5.1 Using performance error metrics

Several statistical performance metrics can be used to measure the performance of the estimated power curves such as the root-mean-squared error (RMSE), normalised mean absolute percentage error (NMAPE), symmetric mean absolute percentage error (sMAPE), the mean absolute error (MAE), and the coefficient of determination (R^2).

In this study, three goodness-of-fit indicators, namely the mean absolute error (MAE), the root mean squared error (RMSE), and the coefficient of determination (R^2) are used to evaluate the goodness-of-fit statistics of the advanced nonparametric power curve models and these are briefly described in chapter 3. Table 4.2 summarizes the error metrics (RMSE, MAE and R^2) for advanced nonparametric power curve models. The RMSE and MAE of the power prediction using the power curve derived from GP power data have lower values than SVM and RF, which concludes the GP model has a better estimation of power curve. While the higher value of R^2 in GP power curve model indicates a better coincidence of observed and estimated results. Based on these three performance metrics, GP algorithm ranks 1 and gives the most accurate power curve while RF based power curve ranks 3 and relatively gives inaccurate power curve.

MODEL	RMSE	R^2	MAE	RANK
SVM	65.086	0.989	46.226	2
GP	62.690	0.990	39.806	1
RF	65.444	0.989	42.568	3

Table 4.2: Evaluation of nonparametric models using performance metrics

4.5.2 Using models residuals analysis

The GP, RF, and SVM models are data-driven, nonlinear techniques whose residual distribution needs to be analysed. Residuals are the difference between measured value and estimated values and can be useful in identifying the deviation between the data and the regression model, which widely used to measure the variability in the response variable. The frequency distribution of the calculated residuals of advanced nonparametric models is shown in Figure 4.10 together with a fitted Gaussian distribution and found that distribution of GP residuals is close to Gaussian as compare to other nonparametric models.

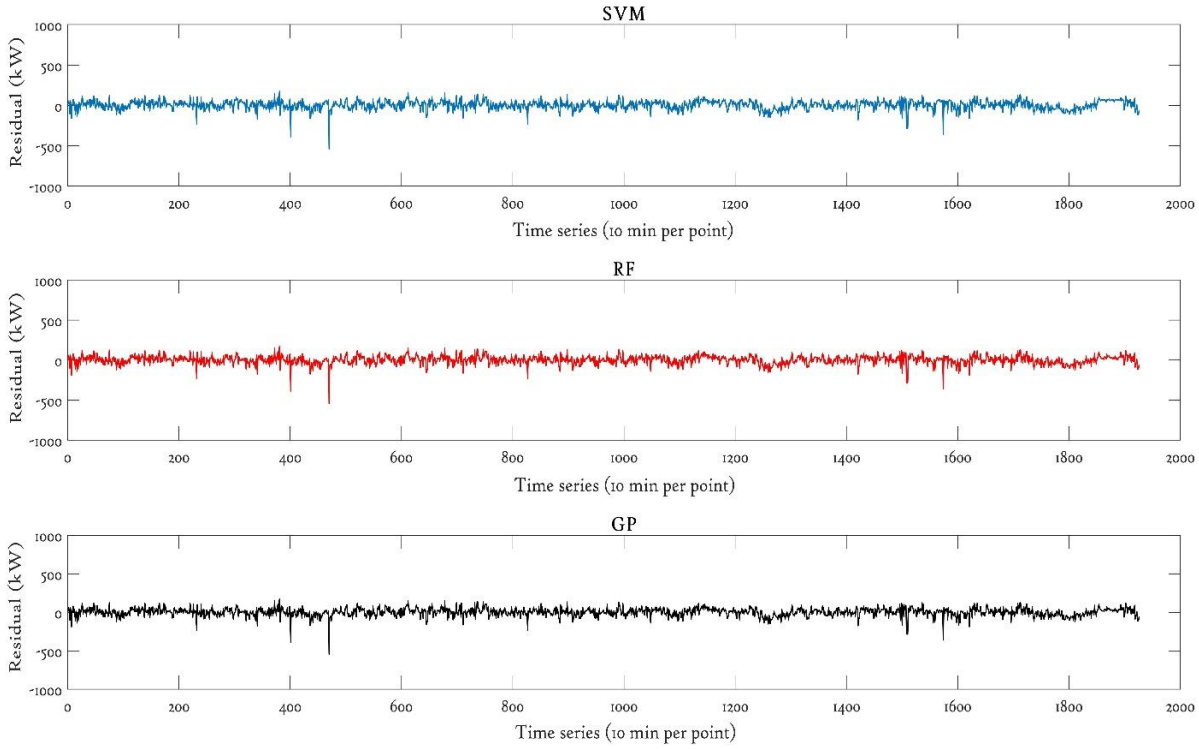


Figure 4.10: Estimated residuals of advanced nonparametric models

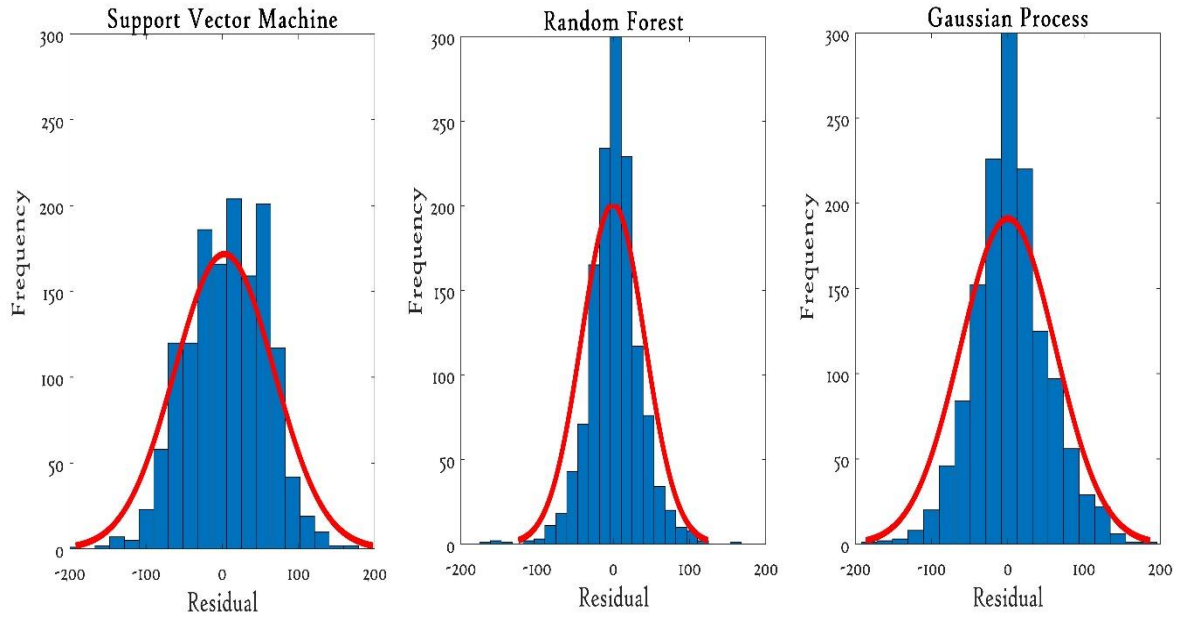


Figure 4.11: Comparative studies of histogram fitting of nonparametric models

4.5.3 Using models uncertainty analysis

Wind turbine power curve vastly used by wind industries to identify the failures that cause the turbine to underperform and do preventive maintenance in order to prevent downtime and catastrophic stage. The GP estimate Confidence Intervals (CIs) along with the mean function which makes uncertainty analysis is simple and straightforward. The detailed methodology to calculate the CIs of GP are described in chapter 3. The dataset of wind turbine asses in context to CIs and if data point lies outside the GP CIs, then that likely to suggest a possible wind turbine underperformance; this is discussed briefly in chapter 7. However, uncertainty analysis in RF and SVM models complex due to the extra mathematical computation associated with it. Some authors proposed techniques to calculate the confidence intervals for RF [133] and SVM [134] models but that requires high power processing, and computational cost and consequently makes the O&M cost higher.

4.6 Chapter conclusions

Three advanced nonparametric models (GP, SVM, and RF) are proposed to assess their comparative performance in estimating power curve and find out of these which model gives most accurate power curve for performance/ condition monitoring purposes. The uncertainty analysis, performance error metric and residual evolution criterion were used for performance comparison of advanced nonparametric models.

The computational results have demonstrated that the GP perform relatively better able to reflect the dynamic properties of a power curve whose distribution function is close to the Gaussian distribution. The calculated values of error metrics further suggest that GP model has the highest fitting accuracy which shows better coincidence of measured and estimated values, followed by the SVM model, see Table 4.2 and Figure 4.8. Both GP and SVM gives continuous and smooth power curves, but the RF based power curve is neither continuous nor smooth because its principle is inspired by CART and weak learners.

The uncertainty analysis is significant in construction robust fault detection algorithms, and CIs comes with estimated values in GP and thus does not require extra mathematical computation which makes fault detection model construction relatively less computational. The uncertainty analysis in RF and SVM requires extra mathematical complexity that ultimately increases the computational cost, and therefore, these models are less attractive for wind turbine condition and performance monitoring activities from economic as well as technical point of view as compared to GP.

In future chapters, how CIs can be useful for GP model uncertainty analysis and thus play a significant role in constructing robust GP fault detections algorithms will be described in detail.

4.7 Chapter references

110. Khalfallah, M., Koliub, A,: 'Suggestions for improving wind turbines power curves'. Desalination, 2007, 209, (3), pp. 221–229. doi: 10.1016/j.desal.2007.04.031
111. R. Chedid, H. Akiki, and S. Rahman, "A decision support technique for the design of hybrid solar-wind power systems," IEEE Trans. Energy Convers., vol. 13, no. 1, pp. 76–83, Mar. 1998.
112. Lydia, M., Kumar, S., Selvakumar,A., Kumar, A.,: ' A comprehensive review on wind turbine power curve modeling techniques'. Renew Sustain Energy Review, 2014, 30, pp. 452-460. doi: doi.org/10.1016/j.rser.2013.10.030.
113. M. Marčiukaitis, I. Žutautaitė, L. Martišauskas, B. Jokšas, G. Gecevičius, A. Sfetsos. Non-linear regression model for wind turbine power curve. Renew Energy, 113 (2017), pp. 732-741
114. Shokrzadeh, S., Jozani, M., Bibeau, E.,: 'Wind Turbine Power Curve Modeling Using Advanced Parametric and Nonparametric Methods'. IEEE Transactions on Sustainable Energy, 2014, 5, (4), pp. 1262-1269. doi: 10.1109/TSTE.2014.2345059.
115. Jung, S., Kwon, S.,: 'Weighted error functions in artificial neural networks for improved wind energy potential estimation'. Applied Energy, 2013, 111, pp. 778-790. doi: doi.org/10.1016/j.apenergy.2013.05.060.
116. Marvuglia, A., Messineo, A.,: 'Monitoring of wind farms' power curves using machine learning techniques'. Applied Energy, 2012, 98, (0), pp. 574–583. doi: doi.org/10.1016/j.apenergy.2012.04.037.
117. Zhang, Y., Wang, J.,: 'K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting'. International Journal of Forecasting,2016, 32, (3), pp. 1074-1080. doi: doi.org/10.1016/j.ijforecast.2015.11.006.

118. Meer, D., Widén, J., Munkhammar, J.,: ‘ Review on probabilistic forecasting of photovoltaic power production and electricity consumption’. Renewable and Sustainable Energy Reviews, 2018, 81, (Part 1), pp. 1484-1512. doi: 10.1016/j.rser.2017.05.212.

119. Hong, T., Fan, S.,: ‘Probabilistic electric load forecasting: a tutorial review’. International Journal of Forecasting, 2016, 32, (3), pp. 914-938. doi: 10.1016/j.ijforecast.2015.11.011.

120. Cortes, C., Vapnik, V.,: ‘Support-vector networks’. Machine learning, 1995, 20, (3), pp. 273-297. doi: doi.org/10.1007/BF00994018.

121. Baccarini, L., Silva, V., Menezes, B., Caminhas, W.,: ‘SVM practical industrial application for mechanical faults diagnostic’. Expert Systems with Applications, 2011, 38, (6), pp. 6980-6984. doi: doi.org/10.1016/j.eswa.2010.12.017.

122. Lin, Y., Kruger, U., Zhang, J., Wang, Q., Lamont, L., Chaar, L.,: ‘Seasonal analysis and prediction of wind energy using random forests and arx model structures. IEEE Transactions on Control Systems Technology, 2015, 23, (5), pp.1994-2002. doi: 10.1109/TCST.2015.2389031.

123. Freund, Y.,: ‘Boosting a weak learning algorithm by majority’. Information and Computation, 1995, 121, (2), pp. 256-285. doi: doi.org/10.1006/inco.1995.1136.

124. Paiva, L., Rodrigues, C., Palma, J.: ‘Determining wind turbine power curves based on operating conditions’. Wind Energy, 2014, 17, pp. 1563-1575. doi: doi.org/10.1002/we.1651.

125. Boser, B. E.; Guyon, I. M.; Vapnik, V.,: ‘A training algorithm for optimal margin classifiers’. Proceedings of the fifth annual workshop on Computational learning theory – COLT, 1992, pp. 144. ISBN 089791497X. doi:10.1145/130385.130401.

126. Fan, R.E., Chen, P., Lin. C.,: 'A Study on SMO-Type Decomposition Methods for Support Vector Machines'. IEEE Transactions on Neural Networks, 2006, 17, (4), pp. 893–908, doi: 10.1109/TNN.2006.875973.
127. Cherkassky, V., Ma, Y.,: 'Practical selection of SVM parameters and noise estimation for SVM regression'. Neural Networks, 2004, 17, (1), pp 113–126. doi: doi.org/10.1016/S0893-6080(03)00169-2.
128. Fan, R.E. , P.H. Chen, and C.J. Lin. Working Set Selection Using Second Order Information for Training Support Vector Machines. The Journal of machine Learning Research, 6:1871–1918, 2005.
129. Lahouar, A., Slama, J.,: 'Hour-ahead wind power forecast based on random forests'. Renewable Energy, 2017, 109, pp. 529-541. doi: 10.1016/j.renene.2017.03.064.
130. Berk, R.,: 'Statistical learning from a regression perspective'. New York, Springer; 2009.
131. Sammut, C., Webb, G.,: 'Encyclopedia of Machine Learning and Data Mining'. Springer, 2017, Boston, MA.
132. Breiman, L. Machine Learning (2001) 45: 5. https://doi.org/10.1023/A:1010933404324.
133. Wager, S., Hastie, T., Efron, B.,: 'Confidence Intervals for Random Forests: The Jackknife and the Infinitesimal Jackknife'. Journal of Machine Learning Research,2014,15, pp. 1625-1651. (accessed on 20th July 2018).
134. Jiang, B., Zhang, X., Cai, T.,: 'Estimating the Confidence Interval for Prediction Errors of Support Vector Machine Classifiers'. Journal of Machine Learning Research, 2008, 9, pp. 521-540. (Accessed on 20th July 2018).

Chapter 5

Wind turbine operational curves using Gaussian Process models

The power curve of a wind turbine is the function that characterises the overall performance of the machine and represents the nonlinear relationship between power output and wind speed. It plays an essential role in the prediction of the energy that can be captured at a site; to make comparisons among concurrent machines and to monitor the turbine efficiency over years of production. Many nonparametric models have been published; mostly applied to power curve based condition monitoring. However, turbine performance cannot be judged solely on power since various additional parameters also have significant influence. With the help of these additional variables, improved nonparametric models can be constructed which may be useful in identifying faults and thus improving wind turbine condition monitoring.

In this chapter, two operational curves, namely, rotor speed curves and blade pitch angle curve along with power curve are constructed using GP approach for performance monitoring of a wind turbine. These developed GP operational curves can be useful for recognising faults that force the turbines to underperform and anticipating failures that will result in downtime.

The constructed GP operational curves are compared with the conventional approach based on binned operational curves and model uncertainty is analysed in order to identify the most effective way to spot operational anomalies.

5.1 Background and Motivation

Unscheduled maintenance resulting from unexpected failures causes downtime and associated loss of generation and potentially increased O&M costs. Several studies suggest that continuous monitoring of a wind turbine can be useful in improving the performance and minimising the O&M cost. The power curve mostly used by wind turbine operator to assess machine power performance, but it is not a perfect indicator because various failures and downtime events remain undetected by it. Therefore, it is desirable to explore other curves that are based on critical parameters that affect the operation and also sometimes the power production of a wind turbine. Wind turbine operation is affected by external factors such as wind turbulence, wind shear and icing, and internal factors such as oil temperature and lubrication. The events related to internal factors can be analysed and potentially controlled, while for external factors this is not possible since they cannot be controlled. Nevertheless, these internal factors are helpful in a performance evaluation of a wind turbine. The internal operation of the wind turbines that affect the power production depends on critical variables, in particular, the rotor speed and blade pitch angle. Therefore, continuous monitoring of these parameters improves the overall effectiveness of any model used to assess turbine performance [135]. The brief literature review of these critical variables and their importance in improving the performance of wind turbines is outlined below.

The power coefficient is a function of the pitch angle, and thus the blade pitch angle affects the power production of the wind turbine. For example, [136], using Blade Element Momentum Theory (BEMT), studied the impact of critical parameters on wind turbine aerodynamic efficiency. Not surprisingly the study confirmed that the blade pitch angle has a direct impact on the power performance of a wind turbine. The nonlinear relationship between pitch angle and wind speed, called the blade pitch curve, is useful for wind turbine condition monitoring. For

example, Singh [137] considered underperformance due to the misaligned wind vane as a case study. A power curve and blade pitch angle curve were developed to detect the performance change due to the misaligned vane, and the result suggests that the blade pitch curve can detect the performance change while it is unidentified by the power curve. Thus, the use of pitch curve monitoring can be beneficial in identifying abnormal behaviour due to failures (e.g., pitch failures). In another example, the authors of [138] used five data-mining methods, namely, bagging, neural network, CART, kNN and genetic programming, to monitor the performance of blade pitch control. Comparative studies of these different models concluded that the genetic programming algorithm prediction accuracy was the best.

Rotor speed in another critical variable that affects the performance of a wind turbine. The rotor curves are useful for identifying failures; there are two types. The rotor speed curve describes the nonlinear relationship between rotor speed and wind speed. A typical rotor curve is a monotonically increasing function of the wind speed and failures of turbines change its shape [139]. Alternatively, the rotor power curve signifies the nonlinear relationship between rotor speed and power output of a wind turbine and is useful in power performance assessment. It should be noted that there exists an optimal rotor speed for a given wind speed for which the power production of wind turbines is maximised. The authors of [140] constructed a reference rotor speed curve on which a multivariate outlier detection approach was applied using k-means clustering and the Mahalanobis distance. Using this curve, the underperformance of a wind turbine was identified using calculated kurtosis and skewness values.

Singh [137] used a rotor curve to compare rotor torque to the produced power, and if there is a difference between theoretical and calculated values, this needs to be investigated. Finally, the authors of [140] studied lookup tables of power-

speed curves used to achieve the maximum power point tracking (MPPT), though to do so requires significant memory space.

5.2 Chapter novel contributions

As already described above, many papers have used the wind turbine power curve to identify abnormal turbine states. However, many failures associated with underperformance and downtime remain undetected by power curve analysis, see for example [137]. Therefore, as substantiated above, there is a need to develop other reference curves based on key performance parameters of the wind turbine such as pitch angle and rotor speed. In this chapter, a SCADA based GP model based on these key variables is presented which can be useful in identifying abnormal operational states and associated anomalies efficiently. Using GP operational curves, a qualitative understanding of turbine health can be used to detect faults at an early stage. Furthermore, the operational curves can be used as performance indicators to measure the impact of internal factors. Moreover, GP operational curves can also be used as a reference model to identify significant wind turbine failures. The IEC 61400-12, [141], uses a data reduction approach, called ‘the method of bins’ or ‘binning’ to calculate the power curve of a wind turbine where ten-minute averaged SCADA data are grouped into wind speed intervals of 0.5 m/sec. The ‘method of bins’ technique is extended to calculate the operational curve and then compared with GP operational curves, and the strength and weakness of individual techniques are identified.

5.3 Wind turbine operational curves

The blade pitch angle, rotor speed, and rotor power were used to construct the GP reference models. The power curve is used alongside these two curves; together they are referred to as the operational curves and described as follows.

5.3.1 Power curves

Wind farm operators widely use the power curve to assess the operation of their wind turbines. Accurate modelling of wind turbine power curve ensures precise performance monitoring and plays a significant role in forecasting wind power generation [135]. A power curve depicts the nonlinear relationship between power output and wind speed and every wind turbine has a characteristic power performance curve. Figure 5.1 is a measured power curve found by field measurements, where an anemometer is placed on a mast reasonably close to the wind turbine. The obtained information from anemometer stored in SCADA system and that information are used to plot the power curve. Figure 5.1 is obtained from monthly SCADA data and using this curve it is possible to predict the energy production of a wind turbine without considering the technical details of its various components. There is always uncertainty associated with a power curve and reducing this is the key for robust performance or condition monitoring of wind turbines. The available active power is a function of the total available power in the wind and depends on hub height wind speed and rated turbine efficiency. Mathematically it expressed as, [134],

$$P = 0.5\,\rho A C_p(\lambda, \beta)\, v^3 \tag{5.1}$$

where ρ is air density (kg/m^3), A is swept area (m^2) , C_p is the power coefficient of the wind turbine and v is the hub height wind speed (m/sec). The shape of power curve is governed by the cubic relation and the way power is regulated not to exceed rated power. Equation (5.1) describes idealized conditions: steady wind; spatially uniform flow undisturbed by any other turbine; no yaw error. Equation (5.1) is referred to as the ideal power curve where the power coefficient depends only on the tip speed ratio (λ) and pitch angle (β). In reality, the power performance of a wind turbine is highly influenced by other parameters associated

with site conditions, for example, wind direction, wind shear, turbulence and others [141].

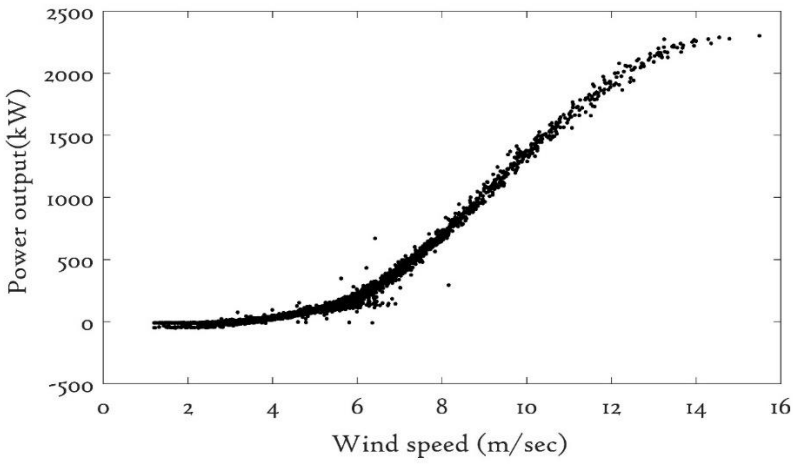


Figure 5.1: Measured power curve.

5.3.2 Blade pitch angle curve

A blade pitch angle curve depicts the relationship between the turbine pitch angle and hub height wind speed, and an example is shown in Figure 5.2. The pitch angle for three bladed wind turbines is either the pitch of an individual blade or is calculated by averaging the angle of all three blades if available. The pitch angle is adjusted by the wind turbine controller to capture maximum power below rated power and to limit power during high winds.

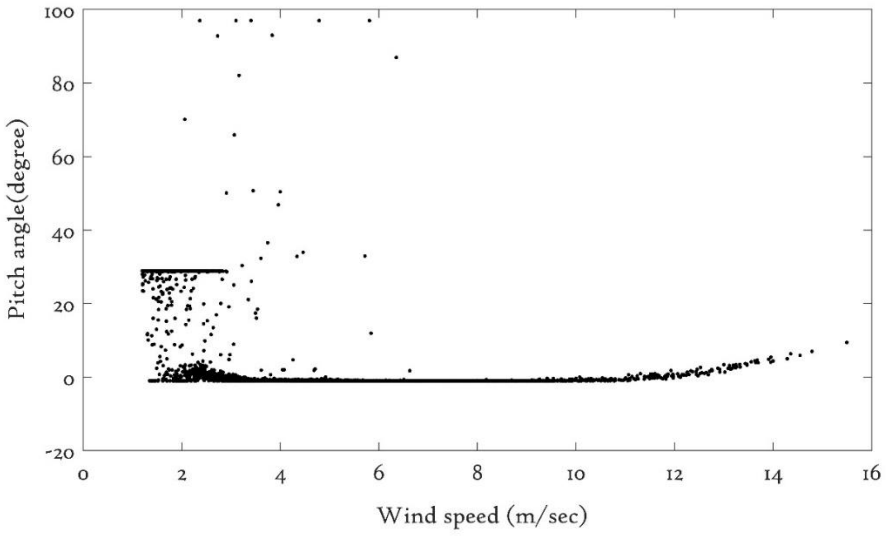
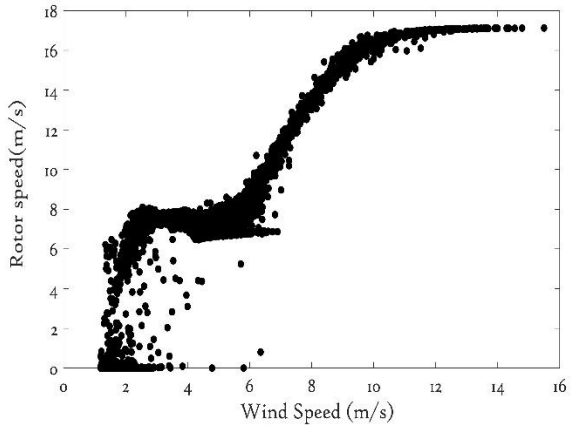


Figure 5.2: Measured pitch angle curve.

5.3.3 Rotor curves

The rotor curves can be classified into two groups: rotor speed curve and rotor power curve. The relationship between rotor speed and wind speed is the rotor speed curve, and an example is shown in Figure 5.3. It can be seen that above the cut in wind speed (4 m/s in this case) the rotor speed increases with wind speed. The rotor power curve that describes the relationship between rotor speed and power output of a wind turbine is illustrated in Figure 5.4. These curves are valuable in identifying faults by means of observing any difference from a reference faulted rotor curve. For example, ref. [137] outlines the importance of rotor power curve in his thesis where the power curve and rotor power curve developed to detect abnormal behaviour of turbines. The comparative studies found that the rotor power curve detected performance changes due to events that remained undetected by the corresponding power curve.



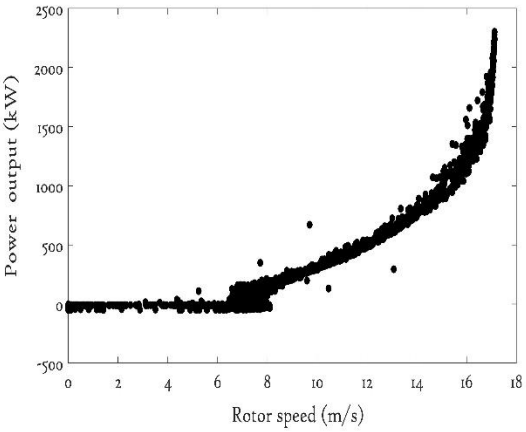


Figure 5.3: Measured rotor speed curve

Figure 5.4: Measured rotor power curve

5.4 Air density correction and SCADA data pre-processing

Power generation from a wind turbine operating below rated power is in theory proportional to air density. Air density is not constant, and it changes with the weather, site altitude, and ambient temperature. Air density is straightforwardly calculated from the measurement of ambient air temperature and pressure. Air

density affects the wind power production, for example, the BARANI Company [142], concludes that weather influences air density significantly and air density can impact wind energy income generation by up to 10%. Air density also affects the accuracy of the power curve, and this is briefly described in chapter 6.

IEC Standard (61400-12-1) suggest two methods for air density correction that can be applied to the power curve depending upon the power control system (i.e., pitch or stall regulated). SCADA datasets used in this thesis are from pitch-regulated wind turbines (WTs) in which a corrected wind speed V_C is calculated using equations (5.2) and (5.3) as shown below,

$$\rho = 1.225 \left[\frac{288.15}{T}\right]\left[\frac{B}{1013.3}\right] \quad (5.2)$$

and,

$$V_C = V_M \left[\frac{\rho}{1.225}\right]^{\frac{1}{3}} \quad (5.3)$$

where, V_C and V_M are the corrected and measured wind speed in m/sec and the corrected air density is calculated by equation (5.2) where B is atmospheric pressure in mbar, and T the ambient temperature in Kelvin in which 10-minute average values obtained from SCADA data are used. The corrected wind speed (V_C) from equation (5.3) is then used to calculate the power curve, normally calculated by binning.

The SCADA data filtration criteria described in section 2.5 are used to remove misleading data and are shown in Figure 5.5 to 5.8. Despite these adopted methodologies, filtered SCADA data are not entirely free from error, but the impact of such errors has been minimized.

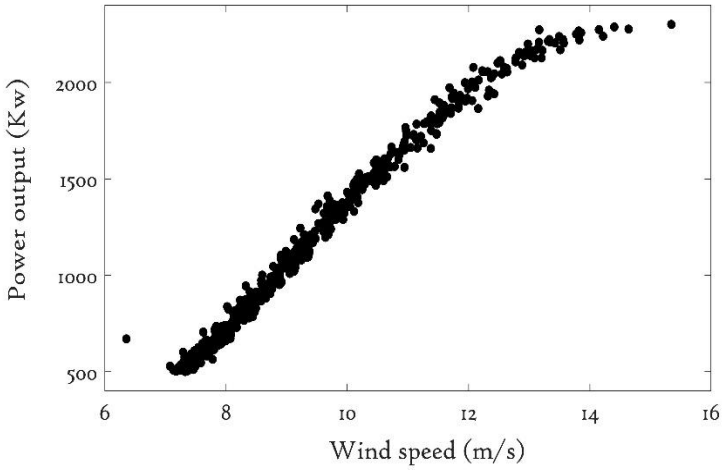


Figure 5.5: Filtered power curve

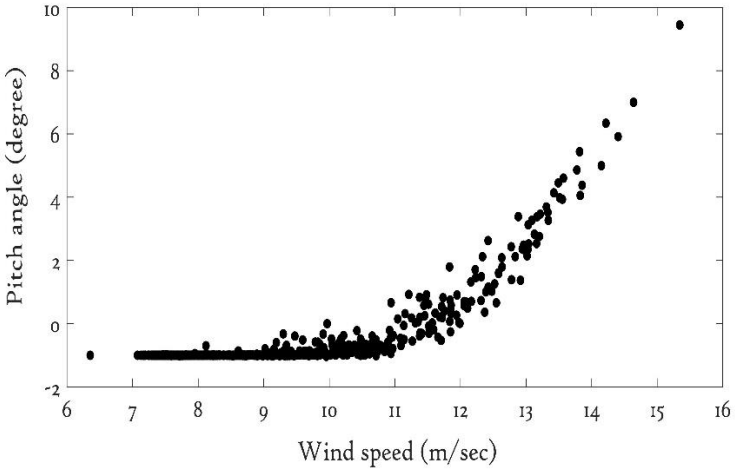


Figure 5.6: Filtered blade pitch curve

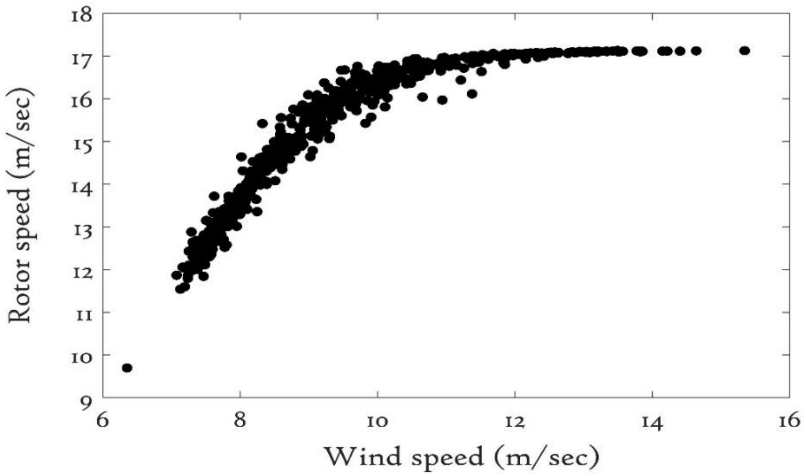


Figure 5.7: Filtered rotor speed curve

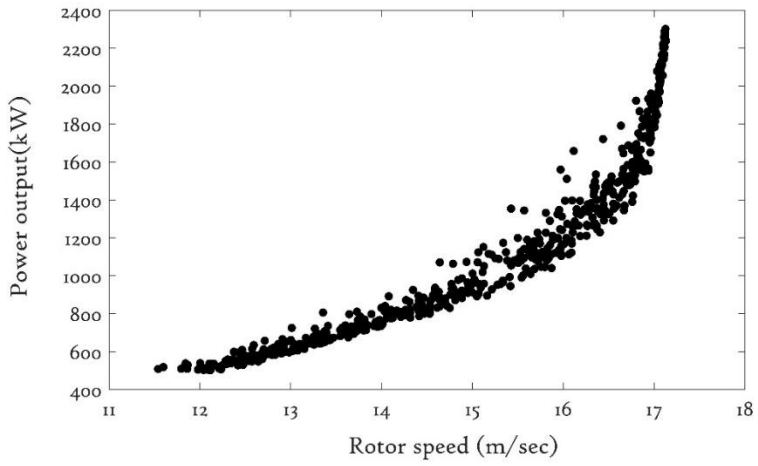


Figure 5.8: Filtered rotor power curve

5.5 Gaussian Process-based wind turbine operational curves

The GP theory described in chapter 3 is used here (realized in MATLAB) to estimate the operational curves of the wind turbine using the filtered, and air density corrected, SCADA datasets (of Figures 5.5 to 5.8), and results are shown in Figures 5.9 to 5.12, with the GP model closely following the expected variance. However, an excessive number of data points can make the GP model inaccurate due to its process involved in inverting a matrix of dimension equal to the number of data points. This asymptotic complexity is called cubic inversion $\boldsymbol{O(N)^3}$ where N is the number of data points. This was outlined in chapter 3.

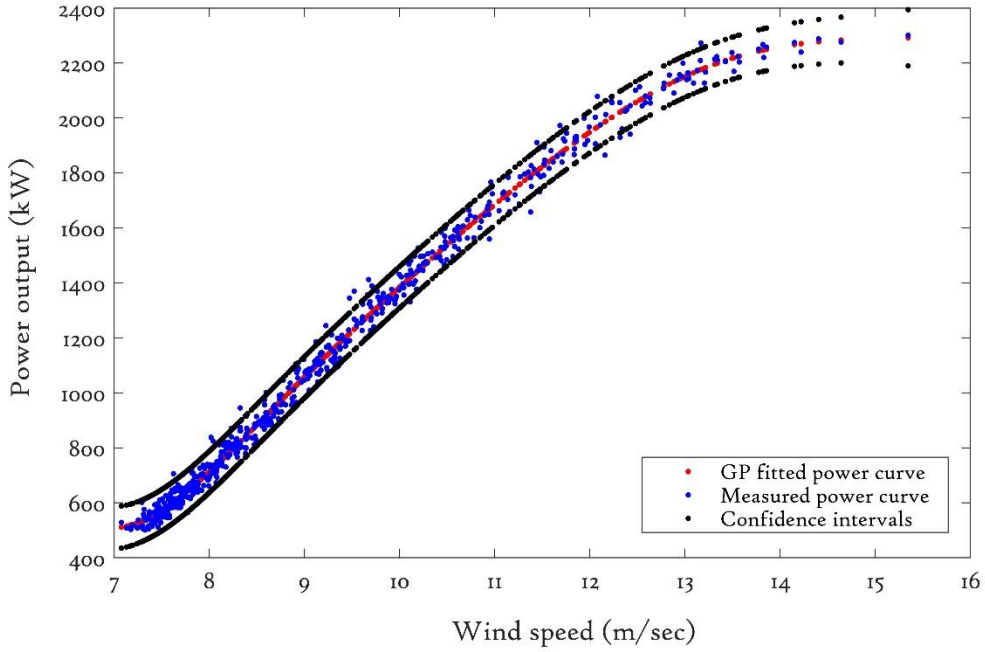


Figure 5.9: GP power curve with CIs

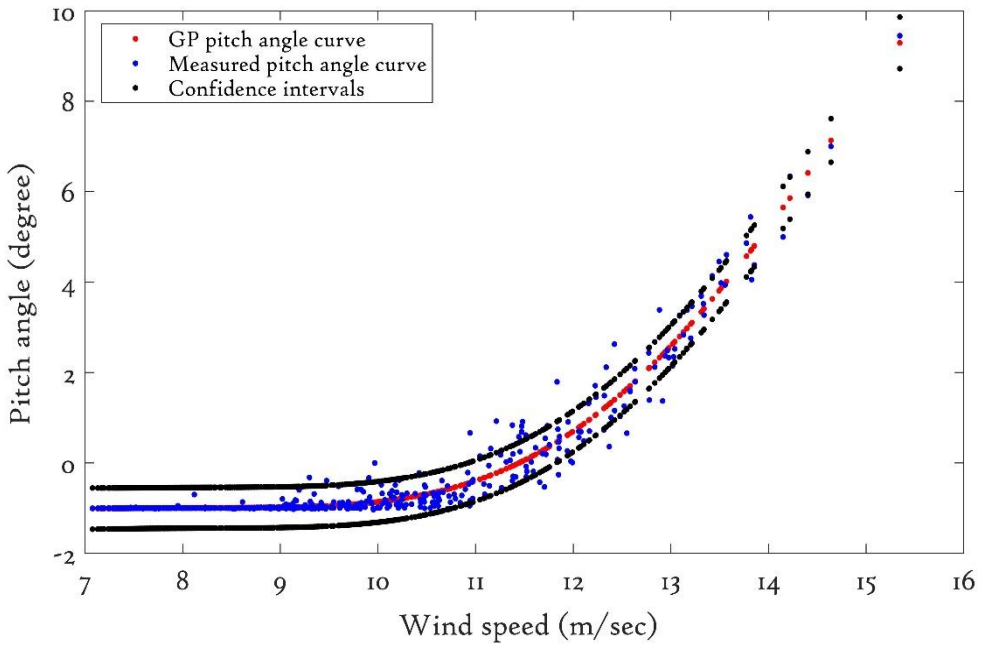


Figure 5.10: GP blade pitch curve with CIs

The black lines of Figure 5.9 to 5.12 show the confidence intervals (CIs) which play an essential role in identifying the unexpected data reflecting operational faults.

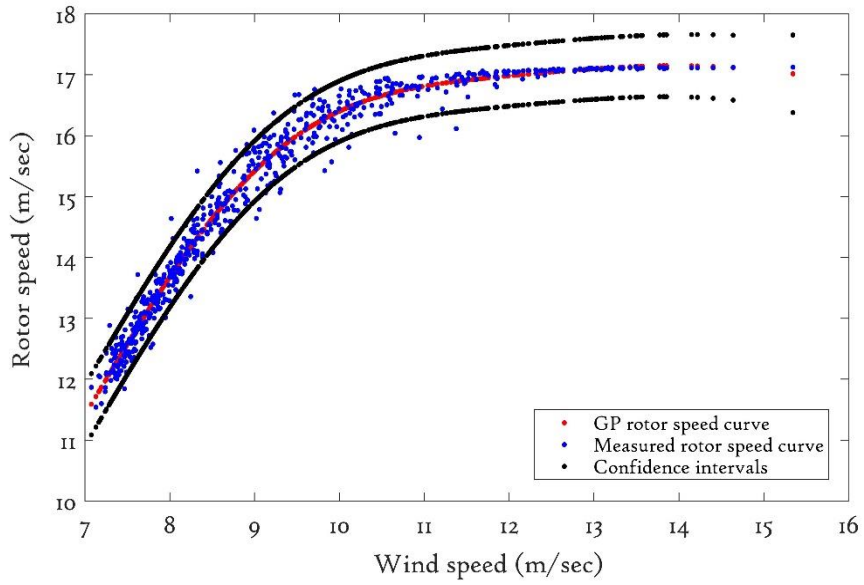


Figure 5.11: GP rotor speed curve with CIs

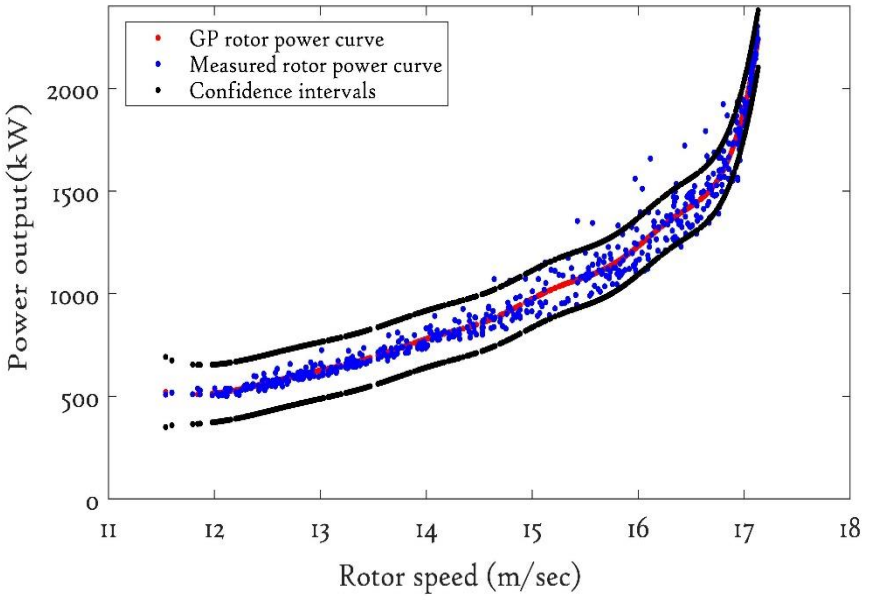


Figure 5.12: GP rotor power curve with CIs

It is worthwhile to note that in figure 5.12 that the CIs is large width from 12 m/sec till 15 m/sec and in Figure 5.11, this is from 10 m/sec to 15 m/sec. While in Figure 5.9 and 5.10 the CIs are not large. While modelling GP operational curve, noise assumed to be constant through the range, plus due to the lack of sufficient number of data points, GP squared covariance function not able to build strong similarity characteristic which can be clearly seen in Figure 5.11

and 5.12 while 5.9 and 5.10 also been affected by this but its impact is relatively small. GP power curve with constant noise assumption works perfectly, and since this thesis solely focused on power curve, hence we continued to use this assumption. However, in dealing with other GP operation curve, it would be better to consider noise point by point basis instead of assuming constant noise across all range in order to get narrow CIs and reduce uncertainty.

5.6 Comparative studies of Gaussian Process based wind turbine operational curves

5.6.1 Using performance error metrics

Deterministic and probabilistic error metrics (RMSE, MAE and R^2) were used to evaluate the performance of GP operational curves. A brief description of the performance error metrics can be found in chapter 3. The calculated error metrics for the three GP operational curves are summarized in Table 5.1 and suggest that a GP accurately estimates the pattern of measured operational curves of a wind turbine. The RMSE and MAE calculated values suggest that among the operational curves, the GP blade pitch curve has a better prediction accuracy while R^2 suggests that the GP power curve distribution function is relatively better than that of other operational curves. Figure 5.13 is the time series comparison of GP operational curves.

GP Operational Curves	RMSE	R^2	MAE
Power curve	38.0403	0.9942	28.9689
Pitch curve	0.2283	0.9653	0.1076
Rotor speed curve	0.2518	0.9773	0.1822
Rotor power curve	69.0489	0.981	46.623

Table 5.1: Statistical measures for GP operational curves models

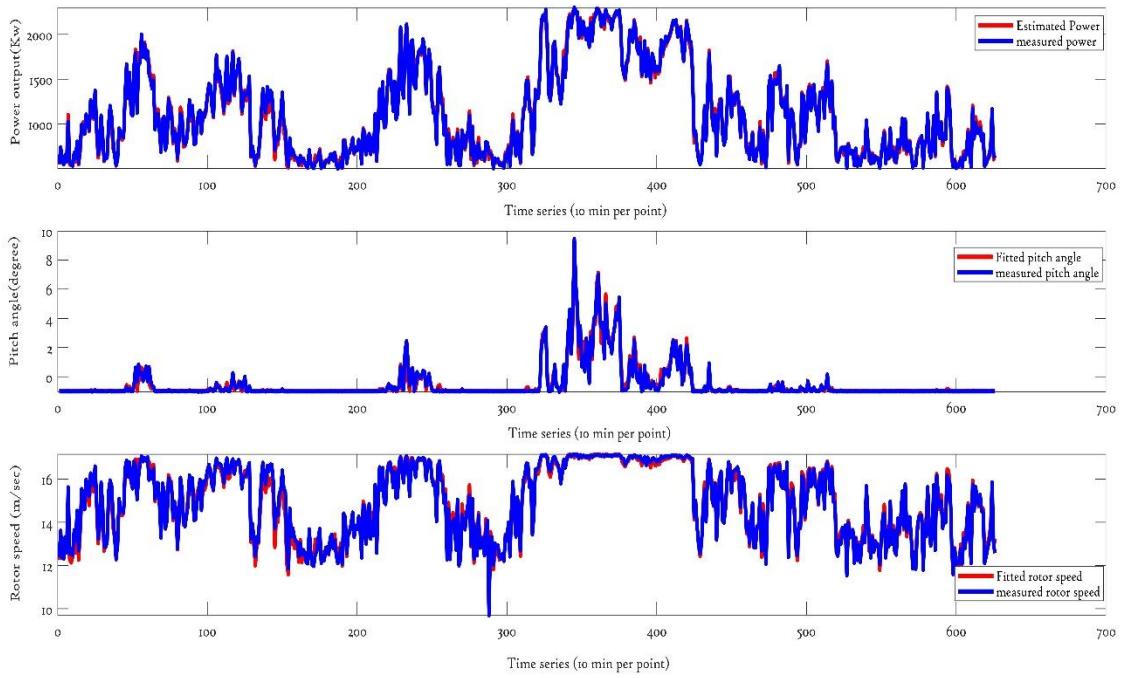


Figure 5.13: Comparison of measured and estimated operational curves in time series

5.6.2 Residual distribution analysis using QQ Plots

The quantile-quantile or QQ plot is a simple graphical technique used to compare collections of data or theoretical distributions and is commonly used to identify the most appropriate distribution function. Identifying whether the distribution function is skewed or slightly tailed can be efficiently analysed with the help of a QQ plot. A theoretical QQ plot examines whether or not a sample S_1,…..,S_n has come from a distribution with a given distribution function F(s) and is plotted against the expected value for the specified distribution using samples of datasets starting from small to large values, [143]. Compared to a histogram, the QQ plot is easy to interpolate. For example, Jean Gibbons [143], indicated that a QQ plot is easier to use than comparing histogram plots in order to judge skewness or more accurately assess whether the distribution tails are thicker or thinner than a normal distribution. Moreover, a QQ plot gives valuable information about graphical properties such as whether shape, location, size, and skewness are similar or different for two distributions and is thus used in this research. Histogram plots based on binning

method have disadvantages compared to QQ plots which is explained in [143] and summarized as follows:

- The binning suppresses the nature and details of the error distribution which is often significant for condition monitoring purposes. For example, binning does not give an accurate view of what's going on in the tails, and also often on in the central section.
- In order to develop an effective binning algorithm, information about bin origin and bin width needs to know since this ultimately affects the appearance of the histogram.
- Comparisons between two histograms are more problematic than that of judging the fit of a group of points to a straight line.

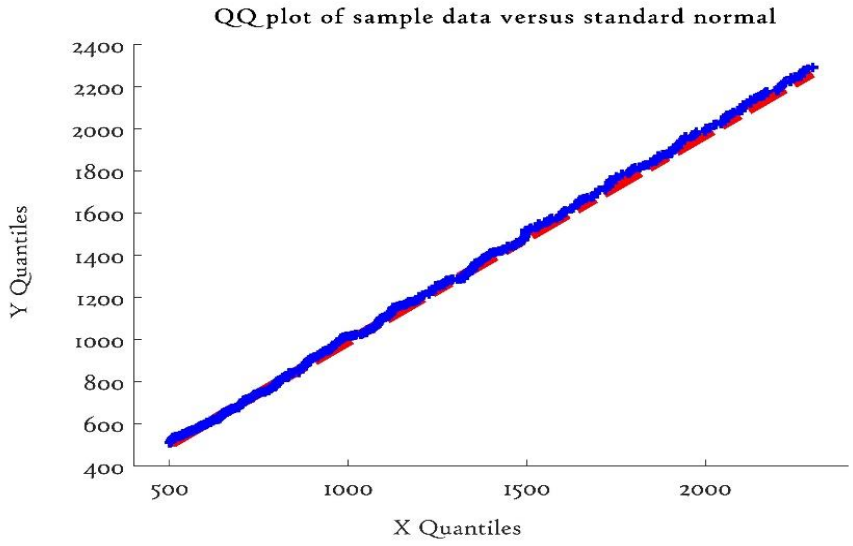


Figure 5.14: QQ plot for GP power curve

As described above, QQ plots that compare two samples of data can be seen as a non-parametric approach to compare their underlying distributions and are hence used here for GP operational curve distribution analysis. Theoretically, residuals of a GP model should be Gaussian, and the typical QQ plot would be a straight line with a unit gradient. It should be noted that QQ plot application is not only limited to Gaussian but can be applied to other distribution function

also. QQ plots are comparing the residual distribution with a Gaussian distribution for GP operational curves (see Figure 5.14 to 5.17). QQ plot plots each data point in ‘x’ using plus sign ('+') markers (blue line) and draws two reference lines that represent the theoretical distribution. A solid reference line (red) connects the first and third quartiles of the data, and a dashed reference line extends the solid line to the ends of the data. This plot produces an approximately straight line, suggesting that the estimated GP operational curves closely follow a normal (Gaussian) distribution. QQ plot analysis is useful for identifying essential parameters such as skewness, associated with GP operational curves. Thus, it is useful in wind turbine condition monitoring.

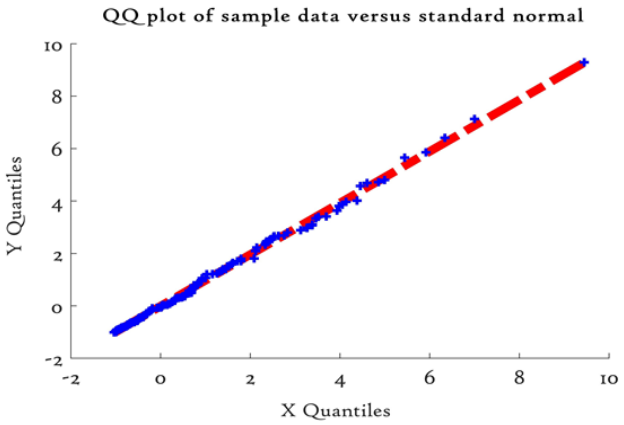


Figure 5.15: QQ plot for GP blade pitch curve

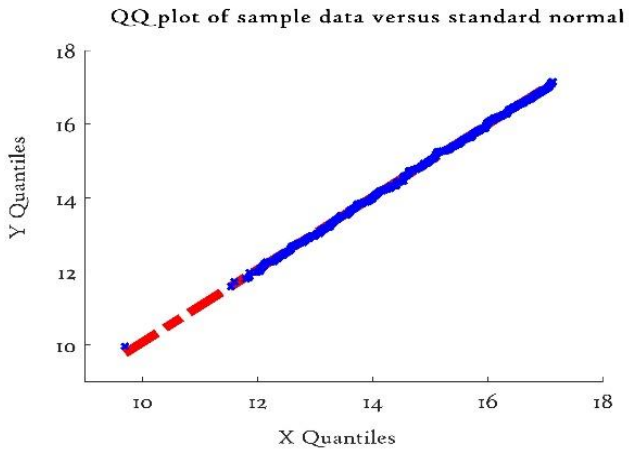


Figure 5.16: QQ plot for GP rotor speed curve

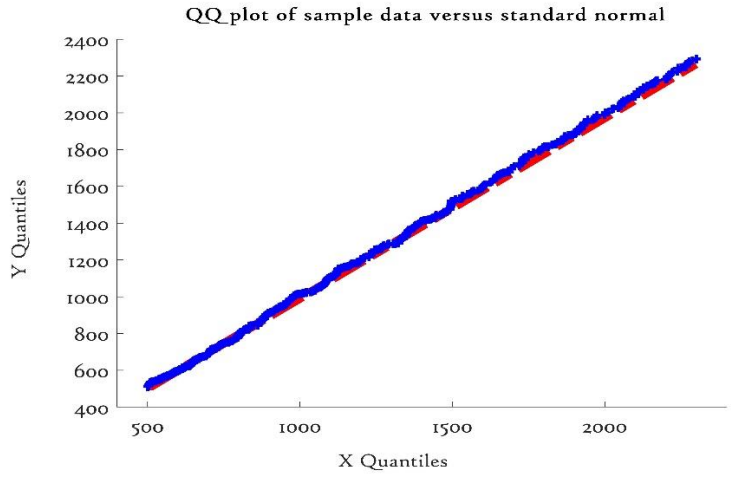


Figure 5.17: QQ plot for GP rotor power curve

Sr.no	GP Operational Curves	RMSE	MAE
1	Power curve	2.22	1.42
2	Blade pitch curve	0.02	0.005
3	Rotor power curve	4.62	1.74
4	Rotor speed curve	0.01	0.004

Table 5.2: Statistical error for different GP models from QQ plot.

Statistical performance indicators (RMSE and MAE) calculated from the QQ plots, further validate this; see Table 5.2. The calculated values of RMSE and MAE suggest that GP-based rotor speed curve and blade pitch curve have a distribution function very close to the Gaussian distribution as compared to other GP operational curves.

5.7 IEC Binning method

The IEC 61400-12 [141] recommended data reduction technique called the 'method of bins' or 'binning' is generally used to calculate the power curve of a wind turbine and its associated uncertainty. The IEC Standard uses mast or nacelle wind speed to calculate the power curve. In IEC binning, the average power output for each bin is obtained by grouping power measurements into wind

speed bins. The bin width kept at 0.5 m/sec wide wind speed interval. Using a large number of data points gives more certainty for the average values in the power curve using this approach. In this section, the binning method has been applied to calculate operational curves of wind turbines using the following equations,

$$V_i = \frac{1}{N_i}\sum_{j=1}^{Ni} V_{n,i,j} \quad (5.4)$$

$$Vr_i = \frac{1}{N_i}\sum_{j=1}^{Ni} Vr_{n,i,j} \quad (5.5)$$

$$P_i = \frac{1}{N_i}\sum_{j=1}^{Ni} P_{n,i,j} \quad (5.6)$$

$$B_i = \frac{1}{N_i}\sum_{j=1}^{Ni} B_{n,i,j} \quad (5.7)$$

where,

V_i = normalised and averaged wind speed in bin i.

$V_{n,i,j}$ = normalized wind speed of data sets j in bin i.

Vr_i = normalised and averaged rotor speed in bin i;

$Vr_{n,i,j}$ = normalized rotor speed of data sets j in bin i.

P_i = normalized and averaged power in bin ;

$P_{n,i,j}$ = normalized power of data set j in bin i.

N_i = number of 10 min average data sets in bin i.

B_i = normalised and averaged pitch angle in bin i.

$B_{n,i,j}$ = normalized pitch angle of data set j in bin i.

It should be noted that the wind speed is the most significant source of uncertainty and including more data points give more certainty to the average value in the pitch curve. Type B uncertainties would be challenging to treat in a consistent manner without greater knowledge of the instrumentation used. Therefore, in this chapter, we used the statistical spread evident in the binned data.

5.8 Comparative analysis of operational curves based on Gaussian Processes and Binning

- **Binned and Gaussian Process based Power Curve**

The IEC binned power curve is considered to be an accurate approach to power curve determination and uncertainty analysis and thus has been selected as the benchmark against which to assess GP models. The error bars are a graphical representation of the variability of wind data and the spread of data around the mean value. They are an effective tool to represent the error or uncertainty in the power curve of a wind turbine. Following the IEC Standard, the binned power curve is calculated as per binning methodology (described above) and is shown with error bars (95% confidence interval) in Figure 5.18. It is found that the spread of data is high (due to high error bars values) between a cut in and rated wind speed, and small above rated, reflecting in the main the manner in which wind turbine power is controlled.

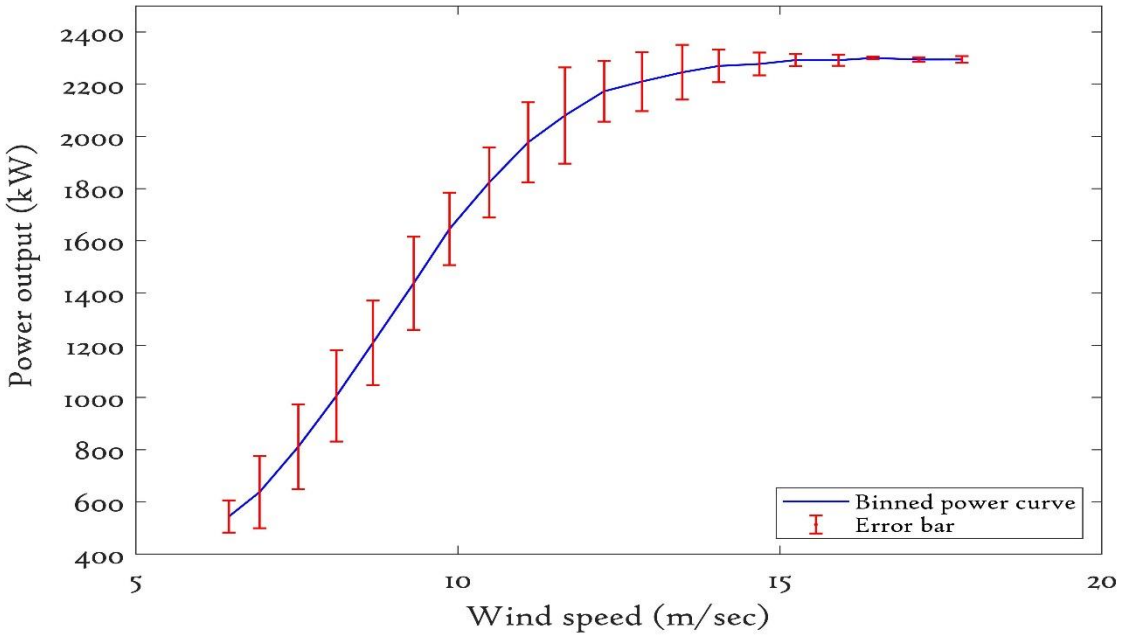


Figure 5.18: Binned power curve with error bars.

Conventional IEC Standard binned power curves described above are compared with a GP power curve shown in Figure 5.19. The GP power curve model closely follows the IEC based power curve. Above rated wind speed, there is fewer

SCADA data available, and as a result, the GP curve is less well determined with some mismatch with the binned power curve. This observation confirms that a GP model based on too little data can be inaccurate. On the other hand, a large dataset leads to high complexity, high processing costs and potentially inaccurate results due to the mathematical challenges posed by the $\boldsymbol{O(N)^3}$ issue. Hence an optimum size of the dataset is a necessary and prerequisite for accurate GP modelling.

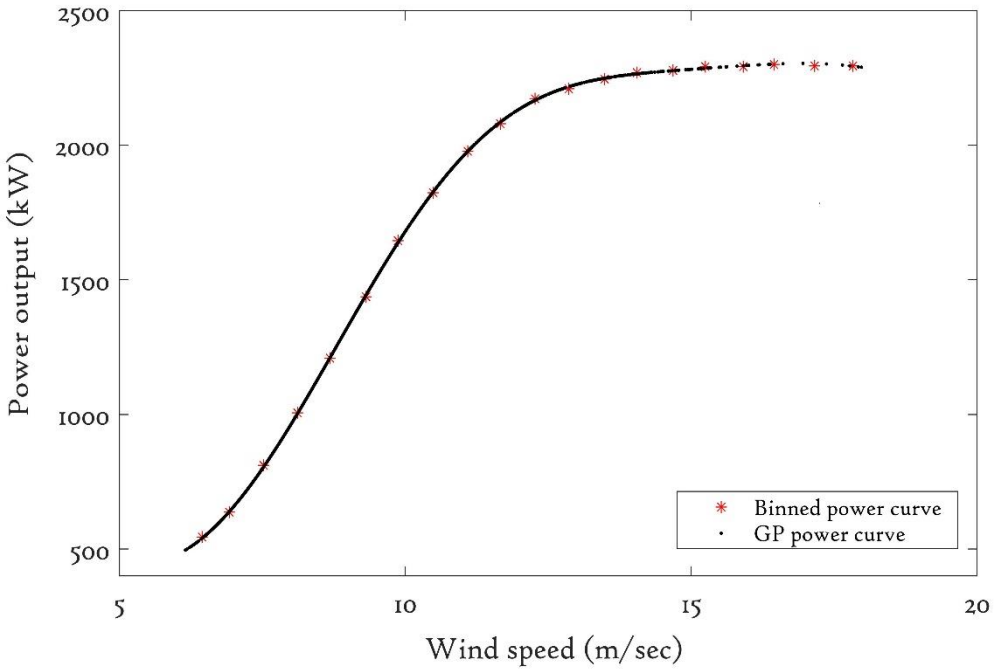


Figure 5.19: Binned and GP power curve comparison.

The GP predicted power curve is compared with the IEC binned power curve (with 95% confidence interval of error bars) in terms of uncertainty in Figure 5.20. The part of the power curve between a cut in and rated wind speeds, where maximum power tracking takes place, is the most critical for condition monitoring purposes. This is because small changes in turbine efficiency (perhaps due to blade damage or excessive drivetrain losses) can be readily detected, whereas, at above-rated wind speed where power is limited by the control system, this loss of efficiency will be masked by greater wind power input. The smaller

CIs for the GP model compared to the binned power curve between a cut in and rated wind speed indicate that GP model is more accurate for this critical range. Since a GP model is a form of interpolation, its uncertainty increases towards the two ends of the data set (i.e., low and high winds). This can be seen in Figure 5.20 but is not of concern for condition monitoring applications for the reasons explained above. It is worth to note that, GP power curve assumed noise constant across wind speed in constructing GP power curve. In short, the CIs of the GP power curve, being smaller for the critical wind speed range, is better able to reject the unhealthy or faulty data than the binned power curve. For accurate and early anomaly detection, the smallest possible confidence intervals are required. The IEC binned power curve is relatively slow to respond while the GP model can be established from limited data with reasonable accuracy. In summary, for a binned power curve, the data variation is highest on the rising section of the power curve. By comparing a binned power curve with a GP power curve, it is found that the latter is more accurate over the rising section of the power curve.

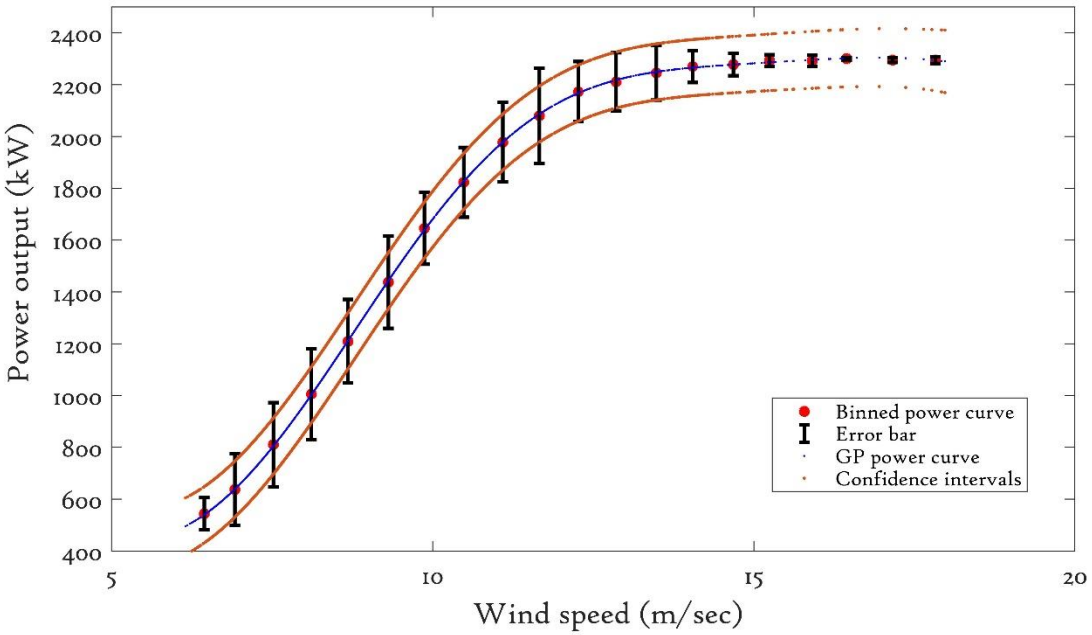


Figure 5.20: Overview of GP power curve and binned power curve uncertainties.

- **Binned and Gaussian Process based blade pitch angle curve**

It should be noted that wind speed is the most significant source of uncertainty and including more data points give more certainty to the average value in the pitch curve. The pitch curve is calculated together with error bars and shown in figure 5.21. It is worth noting that for this study the SCADA datasets used to come from a wind farm located in southern Europe and had an average monthly temperature of 29.779℃. The two standard deviations (i.e., 95% CIs) of measured pitch angle values are used to obtain the error bars which is used to measure the uncertainty associated with each bin of the pitch curve. However, the 'method of bins' is not the most effective technique since its accuracy compromised by choosing bin width of 0.5 m/sec because within each bin the measured power or pitch will depend strongly and non-linearly on wind speed and a wide bin would result in systematic bias. Also, there is a need in practice to get sufficient data points in each bin to be of statistical significance.

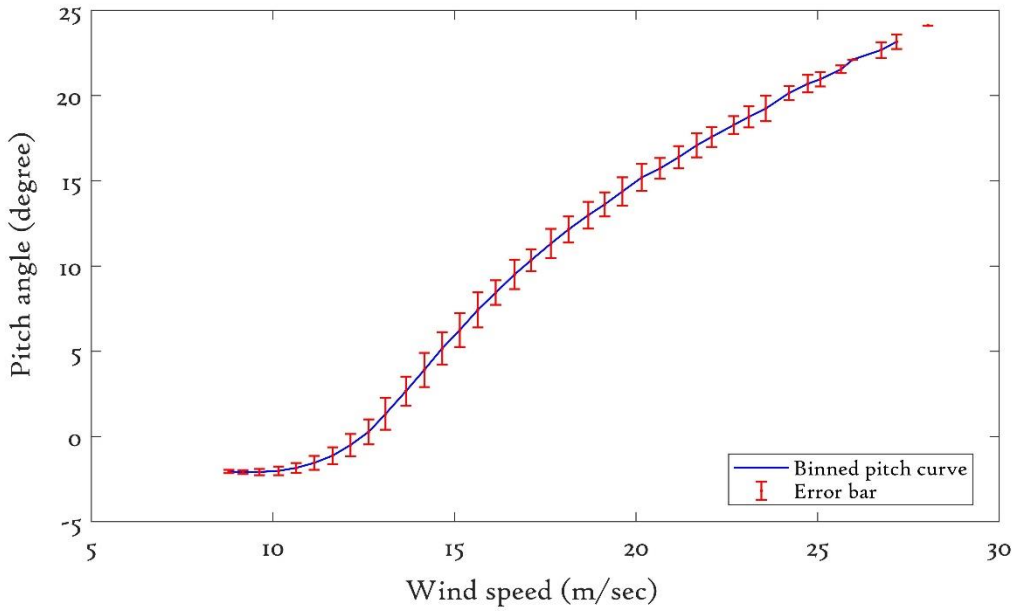


Figure 5.21: Binned blade pitch curve with error bars

Binned pitch curves described above are compared with a GP pitch curve in Figure 5.22; this shows that GP pitch curve model closely follows the binned

pitch curve. This observation confirms that a GP is competent in estimating the wind turbine pitch curve. However, the accuracy of a GP model depends upon the quantity and quality of the data, as well as the appropriate method used. A low number of pitch angle-wind speed pairs may not give a smooth pitch curve while a high number is also not desirable because of the mathematical challenge of calculating a GP for a large number of data points.

A GP model comes with intrinsic CIs, and provide significant information on the uncertainty surrounding an estimation, but they are also model-based estimates, see 3.6.1. The GP estimated pitch curve compared with the binned pitch curve (with 95% CIs of error bars) in terms of uncertainty in Figure 5.23. This comparative analysis suggests that GP confidence intervals are smaller than those from the binned pitch curve in almost all cases. Due to this smaller confidence interval, the GP is better able to reject unhealthy or faulty data than the binned power curve, see Figure 5.23.

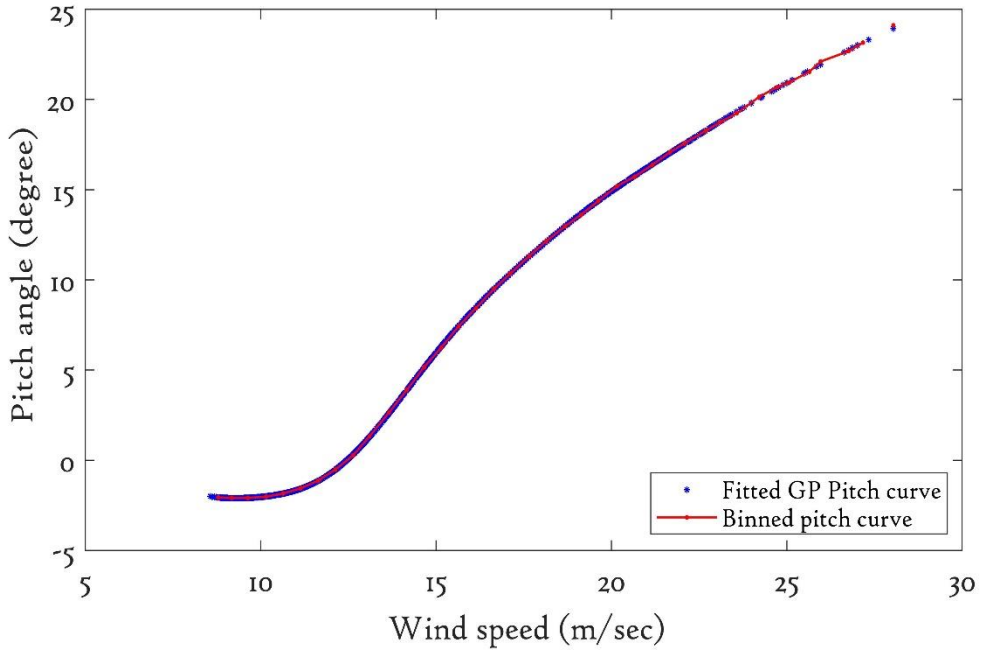


Figure 5.22: Comparison between binned and GP based pitch curve

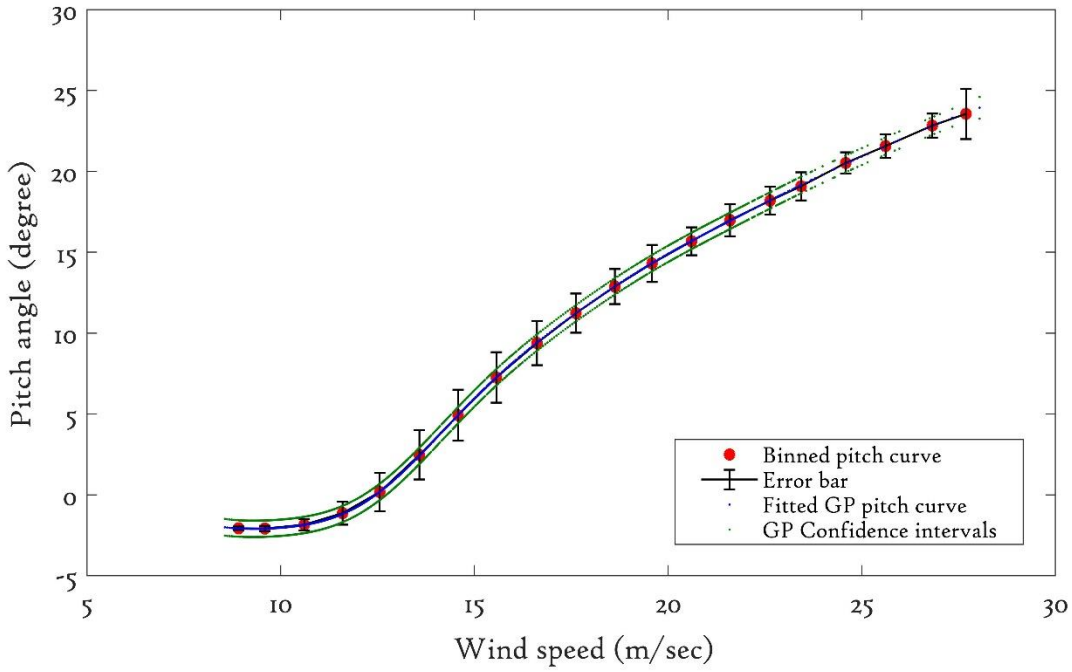


Figure 5.23: Comparative studies of pitch curve uncertainty based on binning and GP

- **Binned and Gaussian Process based rotor curves**

Wind turbine rotor curves constructed together with error bars and compared with measured rotor curves shown in Figure 5.24 and 5.25. The uncertainty of rotor curves based on binning assess via its error bars where two standard deviations (i.e., 95% CIs) of measured power values and measured rotor speed are used to calculate the error bars of the rotor power curve and rotor speed curve respectively. This obtained error bar used to measure the uncertainty associated with a bin of the respective rotor curves. But, the accuracy of 'method of bins' is weaken due to the selection of bin width of 0.5 m/sec because within each bin the output (power, rotor speed) will depend strongly and non-linearly on input (wind speed, rotor speed) and a wide bin would result in a systematic bias, and the need in practice to get sufficient data points in each bin to be of statistical significance. The binned rotor curves as expected following the measured rotor curves of a wind turbine. The uncertainty of binned rotor power curve is started increasing after 16 m/s rotor speed (Figure 5.24) while between cut in and rated wind speed range, the uncertainty of binned rotor speed curve is high, see Figure 5.25.

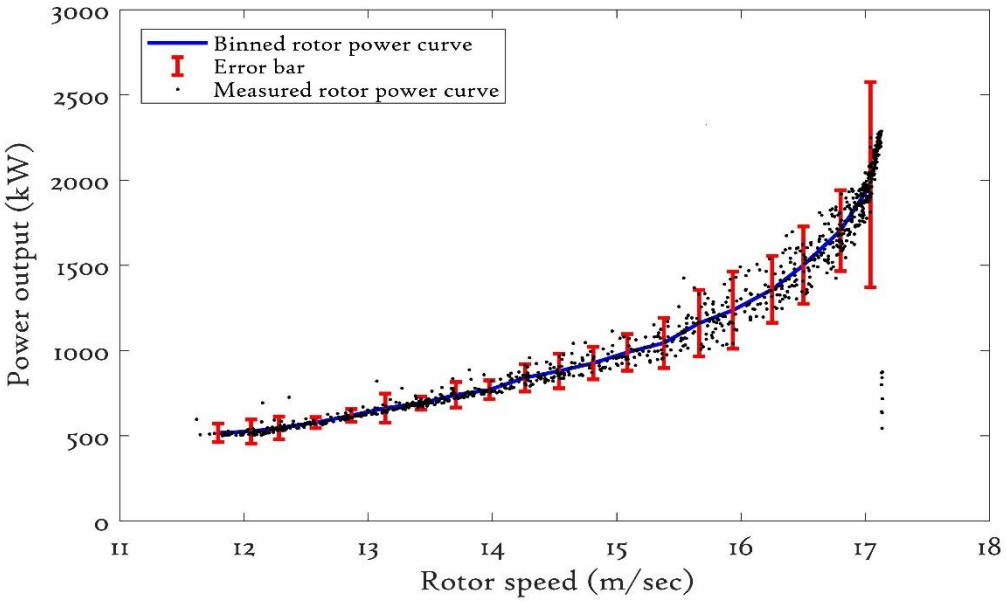


Figure 5.24: Binned rotor power curve

The GP based rotor power curve is able to predict expected variance accurately; this is then compared with the binned rotor power curve and shown in Figure 5.26. However, in this case, the GP model uncertainty as compared to binned approach is relatively high across the rotor speed region.

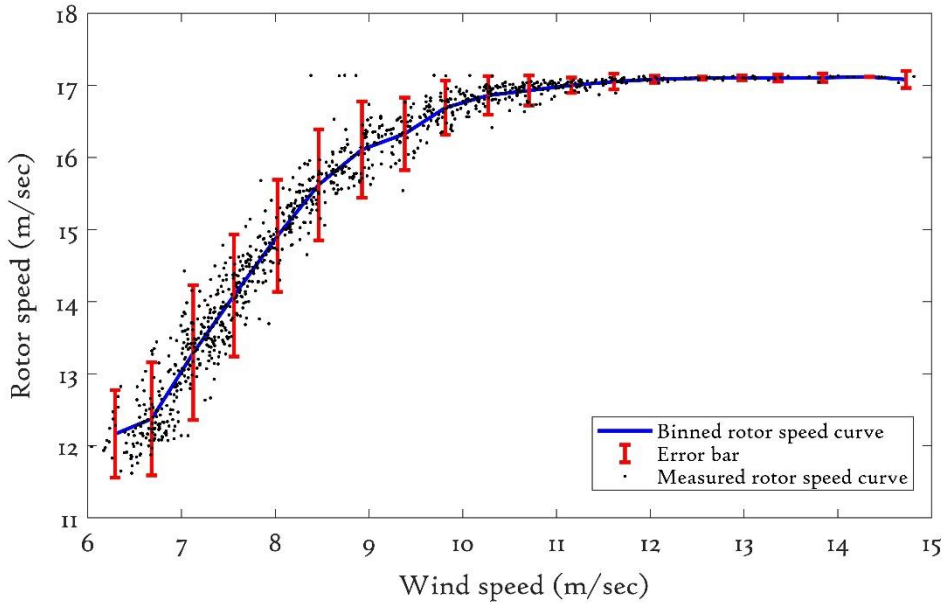


Figure 5.25: Binned Rotor speed curve

The noise in these curves treated constant across the speed range. Although this does not have a significant impact on estimated values, it certainly increases the

GP model uncertainty as shown in Figure 5.26 and 5.27. Therefore, it is advisable to calculate noise value at each point instead of assuming constant across the range in order to get improved GP model uncertainty. Also, it should be noted that the accuracy of GP models depends upon the quantity and quality of the data, as well as the suitable method used.

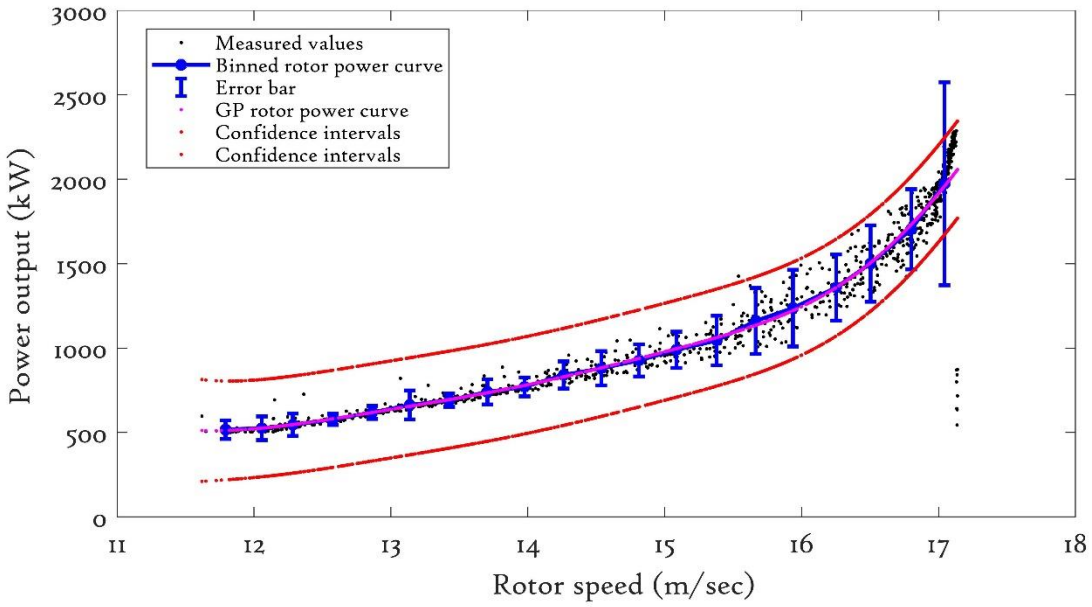


Figure 5.26: Comparative analysis of rotor power curve based on GP and binning

The rotor speed curves based on GP and binning presented and their comparative analysis conclude that the GP model closely follows the binned and measured rotor power curve as shown in Figure 5.27. The uncertainty of the GP model is low between a cut in and rated wind speed range as compared to the binned approach. Above rated wind speed, there are fewer SCADA data available, and as a result, the GP curve less well determined with some mismatch with the binned rotor power curve. This observation concludes that a GP model based on too little data can leads to inaccurate results. On the other hand, a large dataset leads to high complexity, high processing costs and potentially inaccurate results due to the mathematical challenges posed by the $\boldsymbol{O(N)^3}$ an issue associated with cubic matrix inversion (briefly described in chapter 3). Hence an optimum size of the dataset is a necessary and prerequisite for accurate GP modeling.

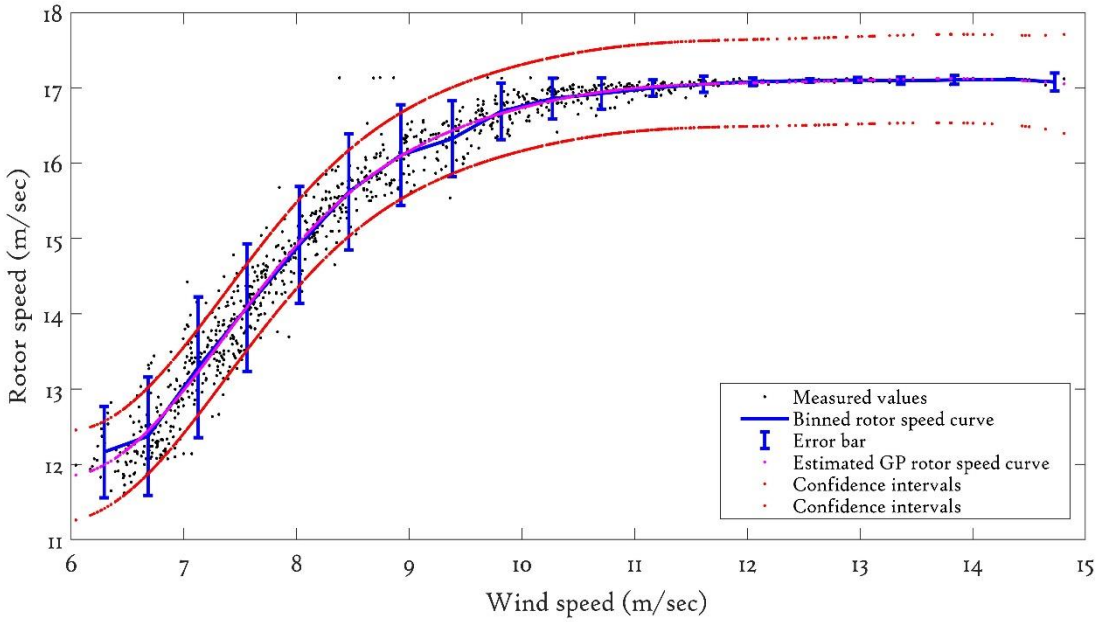


Figure 5.27: Comparative analysis of rotor speed curve based on GP and binning

5.9 Chapter conclusions

The relationship between critical parameters, for example, power, wind speed, blade angle and rotor speed can be used in early detection of faults and failures in order to improve the power performance of wind turbines. GP models are a data-driven approach which is capable of formulating the relationship between these parameters due to its covariance functions and therefore can be helpful in investigating the internal operation of wind turbines. In this chapter, wind turbine operational curves using GP models are presented which can be used to assess underperformance of wind turbines and thus improve their condition monitoring The accuracy of a GP power curve depends upon the quantity and quality of the data but can yield accurate results based on limited data. The estimated operational curves evaluated by performance error metrics, QQ plots, and Residuals analysis; suggest that GP models can estimate the operational curves accurately while QQ plot analysis concludes that GP operational curves distribution functions are close to the Gaussian distribution.

The comparative studies of operational curves based on GP and binning is presented where models uncertainty analysis carried out by CIs (for GP) and error

bar (for binned). For a binned power curve, the data variation is highest on the rising section of the power curve (Figure 5.19). By comparing a binned power curve with a GP power curve, it is found that the latter is more accurate over the rising section of the power curve (see Figure 5.20). While by comparing a binned pitch curve with a GP pitch curve, it is found that the uncertainty is smaller as compared to binned pitch curve in the cut in to cut out wind speed range, see Figure 5.22. The GP rotor curves models uncertainty suffers due to the assumption of constant noise values across the speed range, but in a binned rotor power curve, the data variation is highest on the rising section of the rotor power curve hence perform relatively better, see Figure 5.26 and 5.27. For a binned pitch curve, the data variation is highest between a cut in and cut out wind speed range of the pitch curve (Figure 5.23). Apart from GP rotor curves, treating noise value constant across the range give reasonable accurate results in other operational curves.

In this chapter, operational curves based on the GP model are constructed and its importance in SCADA based wind turbines condition monitoring outlined. The GP operation curves are then compared against binned based operational curve and comparative analysis suggests that GP based operational curve if used as a reference model and passed unhealthy data, fault detection and or performance deviations effectively assessed via uncertainty analysis based on CIs.

In upcoming chapters, a reference operational curve would be undertaken for modelling fault detection algorithm and is then compared against available approaches to judge the effectiveness of operational curve based fault detection models.

5.11 Chapter References

135. Burton, T.; Jenkins, N.; Sharpe, D.; Bossanyi, E. Wind Energy Handbook, 2nd ed.; John Wiley & Sons: Chichester, UK, 2001.
136. Wiratama, I.K.; Nuarsa, I.M. Investigation of factors affecting power curve wind turbine blade. ARPN J Eng. Appl. Sci. 2016, *11*, 5.
137. Singh, P. Analytical Techniques of SCADA Data to Assess Operational Wind Turbine Performance. Master's Thesis, University of Strathclyde, Glasgow, UK, 2013.
138. Kusiak, A.; Verma, A. A Data-Driven Approach for Monitoring Blade Pitch Faults in Wind Turbines. *IEEE Trans. Sustain. Energy* 2011, *2*, 87–96.
139. Kusiak, A.; Verma, A. Monitoring Wind Farms with Performance Curves. IEEE Trans. Sustain. Energy 2013, *4*, 192–199.
140. Li, H.; Shi, K.L.; McLaren, P.G. Neural network based sensorless maximum wind energy capture with compensated power coefficient. IEEE Trans. Ind. Appl. **2005**, *41*, 1548–1556.
141. International Standard IEC61400-12-1. Wind Turbines—Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines; IEC 61400-12-1; British Standard: London, UK, 2006.
142. The BARANI report on Air density changes due to weather 2016. https://barani.biz/kb/air-density-changes-due-to-weather/. 2016. (Accessed August 10, 2018).
143. Gibbons, J.D.; Chakraborti, S. Nonparametric Statistical Inference, 4th ed.; Revised and Expanded; CRC Press: Boca Raton, FL, USA, 2014.

Chapter 6

Incorporating Air density into the Gaussian Process models

Power curves facilitate the prediction of power production at a site and are also useful in identifying the significant changes in turbine performance which can be vital for condition monitoring. However, their accuracy is significantly influenced by changes in air density, mainly when the turbine is operating below rated power [141]. Air density is directly calculated from the measurement of ambient air temperature and pressure. Critical analysis of temperature correction is vital for accurate power curve modelling and application to condition monitoring. Power performance is significantly affected by air density, and thus appropriate investigation needs to be performed on various conceivable air density approaches and their effect on Gaussian Process (GP) power curve modelling. A suitable air density approach selection for GP models not only can improve the power curve fitting but is vital for constructing robust GP power curve models for anomaly detection.

The primary objective of this chapter is to explore whether IEC (a traditional approach) to air density correction is the most effective when estimating power curves using a GP model. To do this definitively, datasets from turbines located in extremely low and high-temperature regions will be used. Since temperature is the most noteworthy factor influencing air density, these datasets have air density values that are very far from the IEC Standard air density, and this is key to understanding its role in GP power curve models.

6.1 Background and Motivation

Power generation from a wind turbine operating below rated power is in theory proportional to air density. The IEC Standard (61400-12-1), [141], references the power curve, known as the standard power curve, to a given standard reference air density. Wind turbine manufacturers supply these standard power curves, and these provide the commercial basis of sales contracts. Accurate wind power forecasting requires accurate power curves and ensures low wind curtailment and underpins efficient wind project planning for construction and operation. Moreover, improved power estimation is also helpful in providing a better prediction of the impact of extreme weather on wind power systems. While constructing a forecasting model, particular attention needs to paid to the uncertainty associated with the forecasts. Jianzhou Wang et al. [144], proposed a nonparametric system to forecast wind speed uncertainty based on recurrence analysis techniques. Based on chaos theory, this approach models the inherent dynamic characteristics of wind speed which assists in exploring the modelling of uncertainty. Also, the same authors, [144], constructed a frequency domain (modal) model to represent uncertainty and results show that this is a more effective and robust than the benchmark models.

Energy produced by a wind turbine is stochastic in nature, hence accurate estimation of any change in conditions is necessary, [145]. Particular attention needs to be paid to the air density since it significantly affects wind power generation and its accuracy. For example, the BARANI Company, [142], concludes that weather influences air density significantly and air density can impact wind energy income generation by up to 10%. Unless compensated for, air density changes will add considerable uncertainty to estimates of long-term energy yield from wind turbines. Air density depends on the specific wind farm locations and most notably on site elevation and ambient temperature, as

demonstrated by [146]. Zahariea and Husaru, [147], confirmed that air density correction (via temperature and pressure) should be implemented when undertaking wind resource assessment and for estimating the performance curve of a wind turbine. The author of [148] has shown that assuming air density to be constant reduces the accuracy of power generation estimates, and so it is advisable to select air density based on the actual wind farm on-site conditions. Here the authors used two wind generators of rated values of 7.5 kW (small-scale) and 850 kW (medium-scale) to analyse the impact of air density on wind generator efficiency. Jung and Kwon, [149], proposed an artificial neural network (ANN) based on error functions to estimate the annual energy production of WTs and results confirmed it is relatively more accurate than conventional ANNs. The authors of [149] utilized three conventional ANN models based on input selections, and these are: a) ANN1 : just wind speed at the reference site used as input; b) ANN2: both wind speeds and directions used as inputs at the reference sites and c) ANN3: used both parameters of ANN2 however in different ways (rather than using the speed and angle, it used *x*- and *y*-components of the wind velocity vector). In Ref. [149], the authors proposed two ANN models; ANN4 and ANN5 in which the inverse of the frequency of the wind speed and power performance curve were used respectively. Both proposed models used the different types of the parameter to calculate the weight applied to each set of training data (C_k). In contrast, the authors of [149] excluded air density variations from their proposed models and argued that its impact in contrast with wind speed variation was limited since energy production is proportional to air density but to the cube of the wind speed. This may be true, but does not support the exclusion of air density from energy yield modelling. Hau, [150], has highlighted that the difference between air densities recorded at sites with a difference in height of hundreds of meters can be such to significantly influence the power performance of a turbine, and thus should not be ignored. Chi Yan and Cristina L. Archer,

[151], carried out an investigation on the effect of compressibility on the performance or large horizontal-axis wind turbines (HAWTs). As part of this study, the authors found that treating air density as constant caused a direct change in power estimation. The work showed that relating compressibility to changing air density slightly degraded the turbine efficiency, confirming the relationship between air density and turbine performance. Liu and Liu, [152], found a significant deviation from standard air density in inland regions of China that are located at a substantial height above sea level; this is further confirm the importance of air density in estimating the power output of a turbine.

Following the IEC Standard, wind turbine power curves are calculated using a data reduction technique known as binning. Air density correction is carried out before binning. However, binning is not necessarily the most effective way to generate a power curve from wind speed and wind turbine power data. As discussed above, many approaches have been proposed that include air density in parametric and non-parametric curve fitting approaches; further examples are [114] and [153].

6.2 Chapter novel contributions

Improving yields from wind farms is vital to operators, and consequently, the need for accurate power curve modelling is increasing. As described above, power performance is affected by air density, and thus appropriate investigation needs to be undertaken to assess the various conceivable air density compensation approaches and their effect on GP power curve modelling.

According to the IEC Standard, as mentioned above, air density correction is required prior to binning. However, there is uncertainty as to whether this approach will also give the most accurate results when using a GP for curve fitting. An important observation regarding air density correction was found in [154], where two different windfarm datasets were used in a GP model. For one of these data sets, the model accuracy marginally improved by avoiding the

standard air density pre-correction. In the present work, data sets with significant air density variations have been explored to confirm the tentative conclusions of [154].

In this chapter different kinds of air density compensation, approaches are explored (including the IEC Standard approach) within a GP model, and their impact on power curve accuracy is assessed. This will lead to an improved air density compensation approach that will prove valuable for the construction of a robust GP algorithm for anomaly detection and other purposes. This the motivation for the work presented here.

To make a definitive assessment, datasets from turbines located in extremely low and high-temperature regions will be used. Since temperature is the most noteworthy factor influencing air density, these datasets have air density values that are very far from the IEC Standard air density, and this is key to understanding its role in GP power curve models. As mentioned above, reference [154] implied that the standard approach to air density correction might not be optimal. The fundamental role of this paper is to demonstrate this definitively for the first time and highlight the preferred approach. It is considered that this has significant implications for all applications of wind turbine power curves and specifically their utilisation in wind turbine condition monitoring.

6.3 IEC Standard air density correction methods

The IEC Standard acknowledges that pressure, temperature, and humidity at the wind farm site affects the air density and so power generation. Out of these parameters, the temperature has the most significant influence on air density and considering its effect has resulted in improved power curves. IEC Standard (61400-12-1), suggest two methods for air density correction can be applied to the power curve depending upon the power control system (i.e., pitch or stall regulated). SCADA datasets used in this paper are from pitch regulated wind

turbines (WTs) and thus following the IEC Standard, the equations (5.2) and (5.3) (of chapter 5) are used to calculate the corrected wind speed V_C,

The air density ρ is related to temperature by the gas law $p = \frac{p}{(R.T)}$; where p is absolute atmospheric pressure, R the gas constant and T is the ambient temperature in Kelvin. The pressure p changes with altitude and can thus vary noticeably according to the location of the wind turbine. The ambient temperature significantly affects the air density and hence power capture of a turbine. Because of the significant influence of temperature on air density, air density correction is sometimes referred to as temperature correction. The power output of the WTs are affected by the weather due to associated air density variations. For a given speed, a WT can produce notably more power in cold weather (due to higher air density) as compared to summer. Henceforth correct air density adjustment is vital for accurate power measurement and prediction forecasting. In power curve analysis, a standard value for air density is used to reflect a typical average air temperature adjusted to sea level.

6.4 Proposed air density correction approaches

As already described above, many published papers have used the IEC Standard air density correction to calculate wind turbine power curves. However, almost no work has been carried on exploring different types of air density correction approach and their impact GP power curve accuracy and uncertainty. This chapter tries to fill this gap by considering four different approaches to compensate (or not) for air density effects when using a GP model rather than binning to identify the power curve. These are:

a) no pre-correction and air density not included in the GP model

b) no pre-correction but with air density included within the GP model

c) pre-correction applied but without air density in the GP model, and

d) with pre-correction and air density included with GP model.

These four approaches are assessed for their impact on the accuracy of the GP power curve model through the confidence limits associated with the fit and by statistical error metrics. SCADA datasets of WTs located in regions with extreme temperatures are used in order to highlight the changes in air density values and the importance of correction.

Figure 6.1 summarises the proposed research methodology in which SCADA datasets featuring extreme temperatures are first filtered and then used to train the GP models. While doing so, four different proposed air density compensation approaches are incorporated into the GP models in order to analyse their impact on GP model accuracy and uncertainty. Confidence intervals, performance error metrics and calculated residuals are used to assess the impact of these four air density compensation schemes on the accuracy of the GP models.

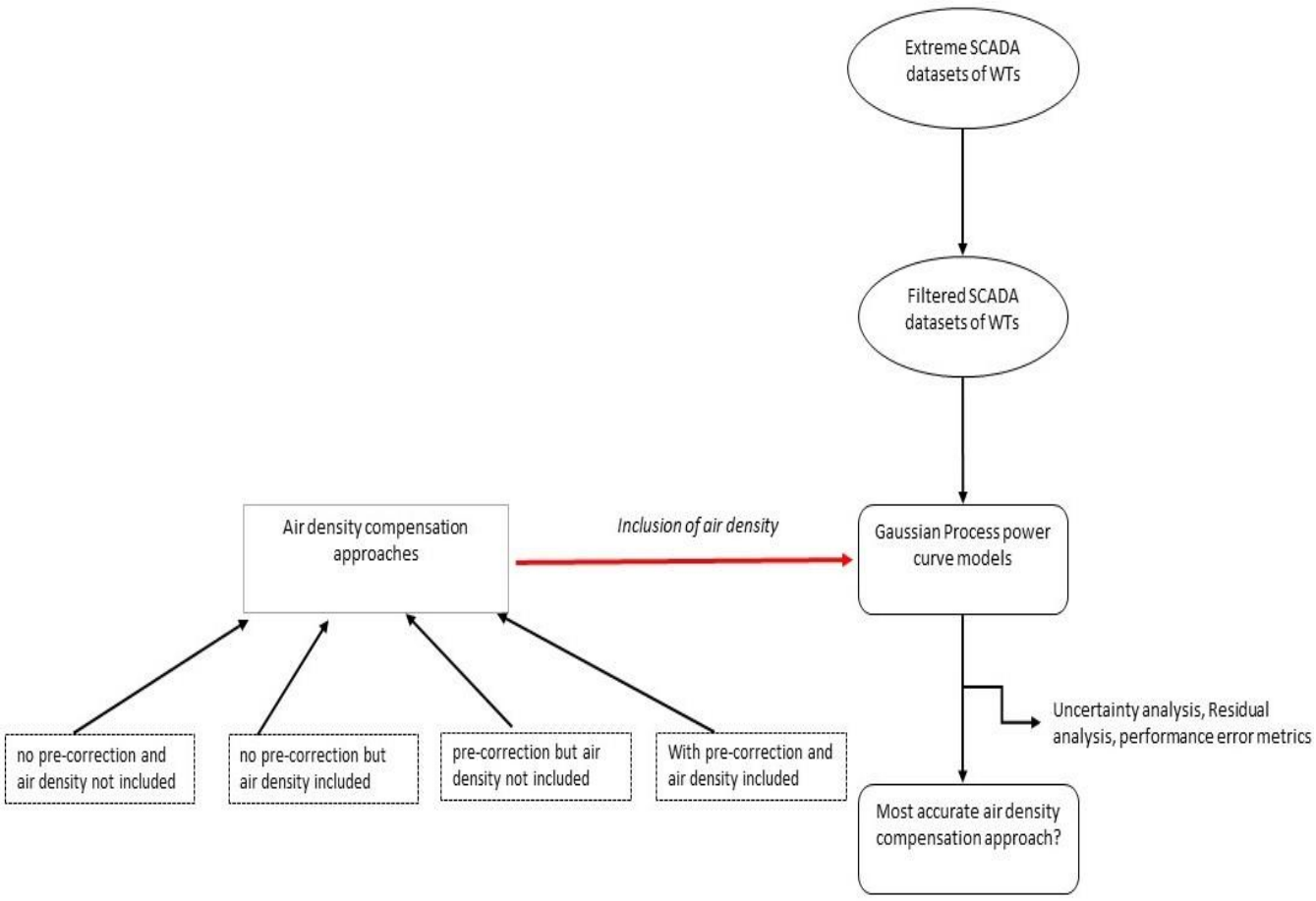


Figure 6.1: Flowchart of air density compensation approaches and GP models.

6.5 Extreme SCADA datasets description

According to the IEC Standard (61400-12-1) guidelines, air density correction shall be applied when the site density differs from the standard value ($1.225\ kg/m^3$) by more than 0.05 kg/m^3. To ensure that, SCADA datasets from turbines located in extremely low (northern Europe) and high-temperature regions (southern Europe) will be used. Here, 'extreme' refers to the SCADA datasets that have the highest and lowest average monthly temperature values. Because temperature is the most noteworthy factor influencing air density. These datasets have air density values that are very far from the IEC Standard air density, and this is key to understanding its role in GP power curve models. For the sake of convenience, the two sites are designated as follows:

A = northern Europe site; B = southern Europe site.

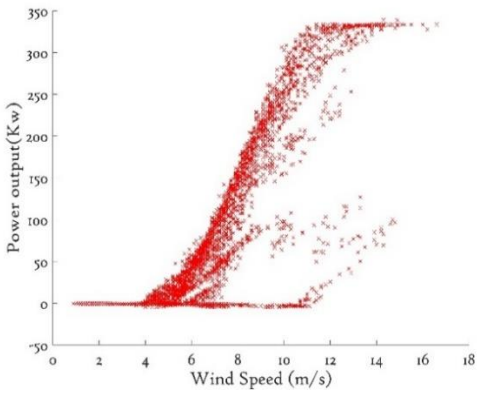
WTs	SCADA datasets time period	Total number of data points	Average monthly temperature (℃)	Standard density ($\boldsymbol{kg/m^3}$)	Mean absolute difference ($\boldsymbol{kg/m^3}$)
A	1/02/2010 - 28/02/2010	4032	-5.2775	1.225	-0.102
B	1/08/2010 - 31/08/2010	4400	29.7791	1.225	0.061

Table 6.1: Description of selected monthly SCADA datasets from A and B.

The datasets used are from the year 2010 for which February and August SCADA data from sites A and B record extreme average monthly temperatures of $-5.277℃$ at A and $29.779℃$ at B. For these two particular months the air density differs by -0.102 and 0.061 from the IEC Standard value for sites A and B respectively as shown in Table 6.1. Thus both these datasets conform to the IEC Standard guideline described above for making air density correction. Note that all SCADA datasets used in this chapter comprise 10-minute averages with

maximum, minimum, standard deviation and an entire month of operational data from the two sites.

Systematically remove SCADA data errors due to sensor failures will affect power curves and should as far as possible at the outset. Pre-processing and air density correction methods described in section 2.5 and 5.4 respectively are used to filter and air density correction. Table 6.1 summarises the datasets of turbine A and B. Set A starts with time stamp ‘‘1/02/2010 00:00 AM’’ and ends at time stamp ‘‘28/02/2010 23:50 PM’’ and contains 4032 measured values which reduced to 1205 data points after pre-processing, while B starts with time stamp ‘‘1/08/2010 00:00 AM’’ and ends at time stamp ‘‘31/08/2010 23:50 PM’’ and contains 4400 which became 2068 data points after pre-processing. The GP models power curve models based on the entire training data and then used for further analysis. Figures 6.2 and 6.4 show the monthly unfiltered power curve data for the turbines at sites A and B respectively. Note that this data has not in any way been adjusted to reflect air temperature. As already mentioned, the mean absolute value of the air density for each of the WT has been calculated (Table 6.1). The differences between the months can be useful in understanding the analysis of power curve fitting. Figures 6.3 and 6.5 are the filtered monthly power curves for the respective turbines after pre-processing.



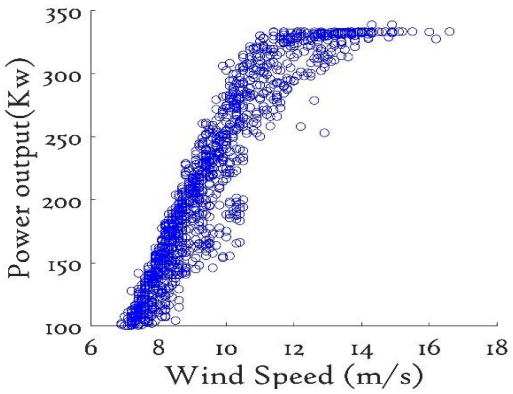
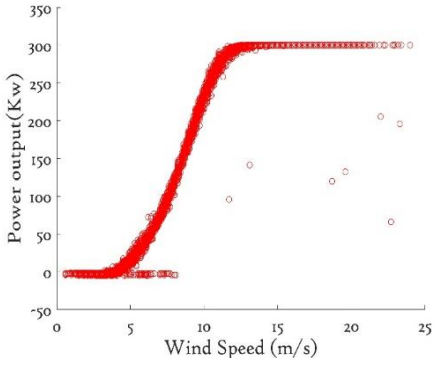


Fig 6.2: Raw data for turbine A dataset Fig 6.3: Filtered & corrected data for turbine A



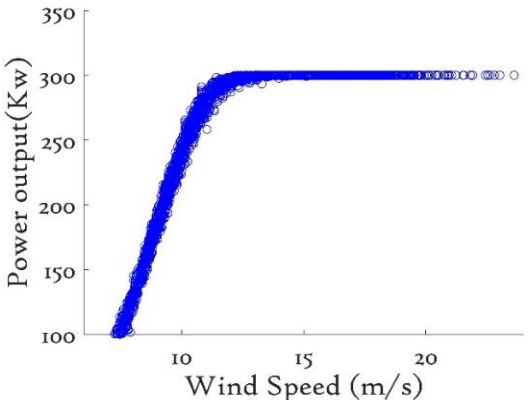


Fig 6.4: Raw data for turbine B dataset Fig 6.5: Filtered & corrected data for turbine B

6.6 Gaussian Process models under different air density compensation approaches

The GP theory described in chapter 3 (realised in MATLAB) are used to model the power curve for the different approaches to air density correction, and results are illustrated in Figures 6.6 and 6.7. In order to evaluate the smoothness of estimate GP power curve model under different air density correction approaches, the standard deviations of the difference are calculated. If the difference of standard deviation for measured values and estimate values are close and small, then it suggests the estimated power curve is smooth and of good fit. For example, standard deviations of the difference are calculated from the estimated GP model (of Figure 6.7) in which 'no pre-correction but density included' considered. The standard deviations of the difference for measured and estimated values are 22.79 and 21.94; both calculated values are close and thus suggest the strong smoothness. The similar approach is being taken for GP power curve model based on different air density compensation approach. Result suggests that the standard deviations of the difference of estimated and measured values are very to each other in GP power curve based air density correction approaches for both extreme SCADA datasets.

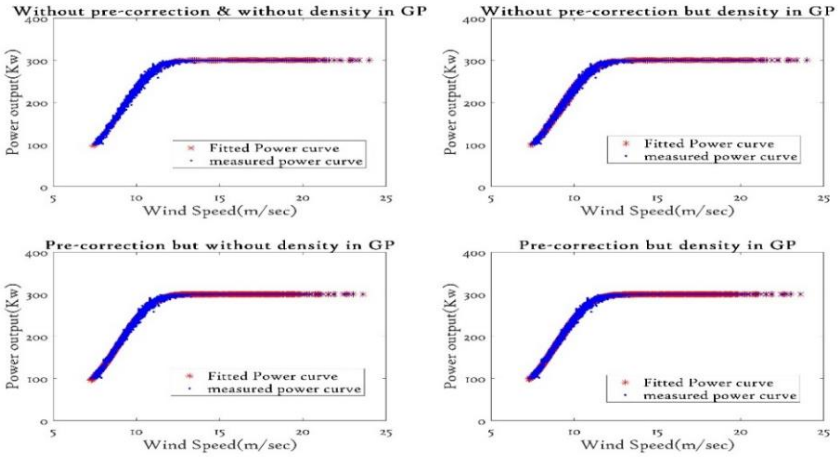


Figure 6.6: GP power curve fitting for different air density approaches using site B data

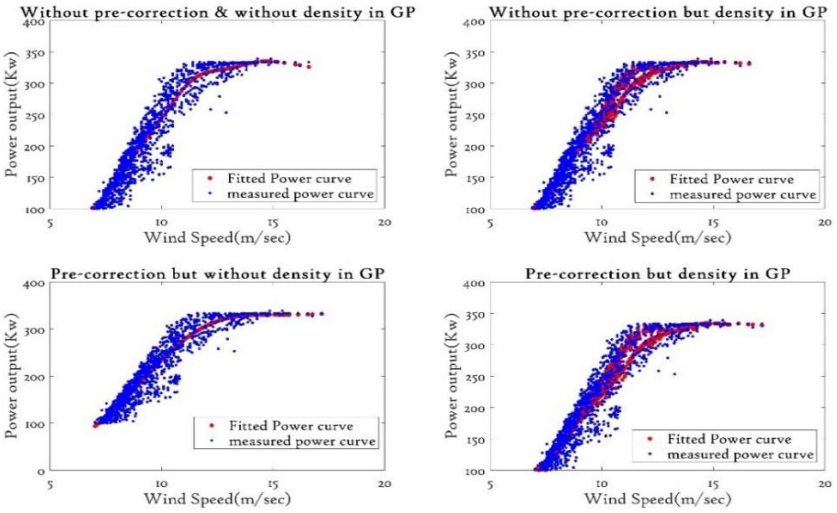


Figure 6.7: GP power curve fitting for different air density approaches using site A data

The SCADA data of WT B has limited scatter and thus estimated GP power is smooth, see Figure 6.6. In contrast, the SCADA dataset of WT A is more scattered (due to the extremely low ambient temperature values) and also has fewer data points at the rising section of the power curve. This issue along with cubic inversion problem (due to the involvement of inverting a matrix of dimension equal to the number of data points) described in chapter 3 affects the GP model fitting accuracy and increases the computation cost, so striking a balance between the number of data points used and computation cost is critical for effective GP

modeling. For these reasons, the data sets will be restricted to a wind speed range of 8 to 14 m/s. For a wind speed range of 8 to 14 m/s, the two data sets have almost the same number of data points, and the data from turbine A is less scattered. It should be noted that the accuracy of a GP power curve depends upon the quantity and quality of the data. A low number of power-wind speed pairs may not give a smooth power curve while a high number is not desirable because of the numerical challenge of calculating the GP.

6.7 Impact of different Air density compensation approaches on GP Power Curve models – A Comparative Study

In this section, the accuracy of the GP power curves with the different approaches to air density compensation using the residual distributions and a range of statistical measures (MAE, MSE, MAPE, RMSE, R^2) are presented . As mentioned above, to limit the size of the data set, the analysis will be restricted to the wind speed range of 8 to 14 m/s since the number of data points are reasonable within this range and also the number of data points resulting for sites A and B are almost the same and less scattered. This has the further advantage that the measurement errors (assumed to be constant in the GP) are reasonably consistent. The mean absolute value of the air density correction, has been recalculated for this restricted wind speed range, and is shown in Table 6.2.

Site	Standard density	Mean absolute density difference	Total number of data points used
A	1.225	0.099	1114
B	1.225	0.062	1116

Table 6.2: Mean absolute value of standard air density correction for restricted range data sets

6.7.1 Uncertainty assessment using confidence intervals

A GP power curve model intrinsically represents fitting errors and thus model accuracy. An accurate assessment of a GP power curve can be done using confidence intervals (CIs) that are useful to portray how individual air density approaches affects the GP model uncertainty. GP power curve fitting comes with intrinsic confidence intervals (that reflect the standard deviation of the model), and thus provides information on the uncertainty surrounding an estimation but are itself a model-based estimate. Data points that lie outside of the confidence intervals can be considered anomalous, signifying a potential malfunction of the WT. Also, uncertainty signifies the 'goodness' of the estimated fit. Confidence intervals are a useful measure of uncertainty and the precision of model estimates. Recall from chapter 3, the confidence intervals of the GP model is calculated by equation (6-3),

$$CIs = \overline{f*} \pm 2\sqrt{Var[f*]} \tag{6-3}$$

In equation (6.3), CIs signifies the pointwise mean plus and minus two times the standard deviation for given input value (corresponding to the 95% confidence region which represents the significance level of 0.05), for the prior and posterior respectively. CIs have been calculated for the restricted data sets for sites A and B and plotted against the wind speed and are shown in Figures 6.8 and 6.9. The result suggests that the traditional approach to air density correction does not yield the most accurate GP model. The most accurate approach is not to undertake air density pre-correction but to include air density within the GP model.

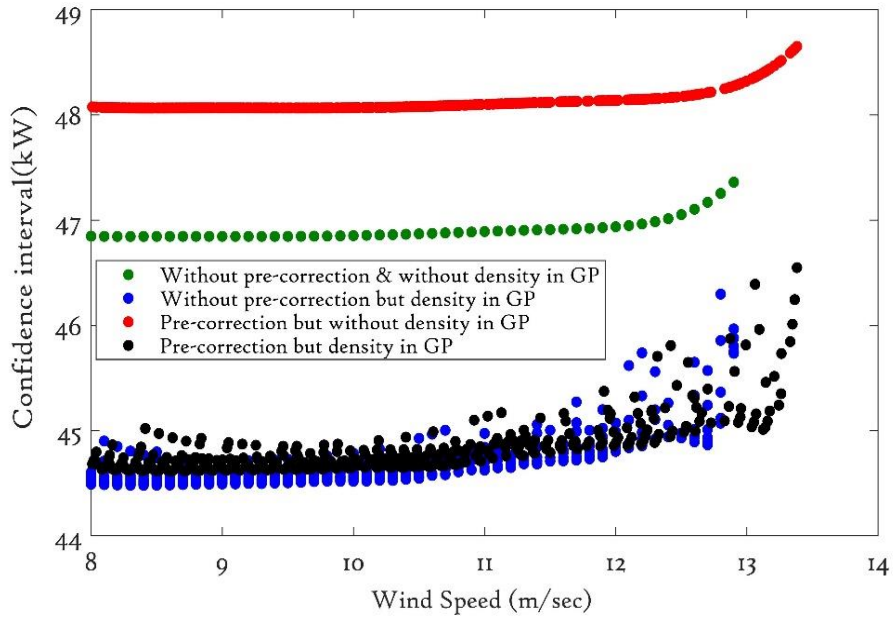


Figure 6.8: Uncertainty assessment in terms of confidence intervals for different air density approaches with limited data set for site A

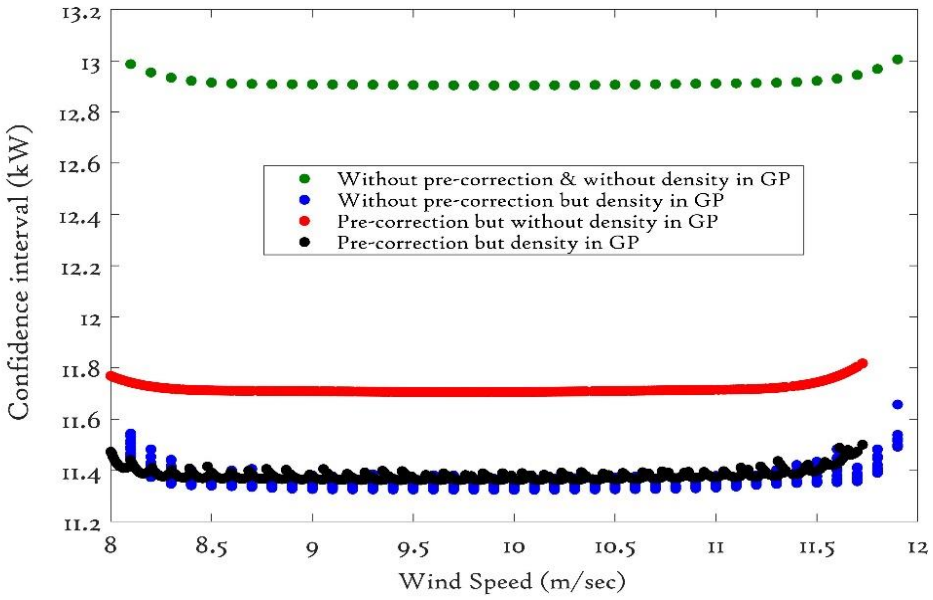


Figure 6.9: Uncertainty assessment in terms of confidence intervals for different air density approaches with limited data set for site B

6.7.2 Performance error metrics

Five performance error metrics (described in chapter 3), namely mean absolute error (MAE), mean square error (MSE), mean absolute percentage error (MAPE), root mean square error (RMSE) and coefficient of determination (R^2) are

calculated. They are used here to assess the accuracy of the GP power curve models for the different air density correction approaches. The results of these statistical measures confirms that adding density into GP model that uses data that have not been pre-corrected for air density gives the most accurate result as compared to other GP models and results as shown in table 6.3. The conclusion is that with a GP model, pre-correction does more harm than good, and this is assumed to be due to the imperfect nature of the pre-correction.

Continuous rank probability score (CRPS) is another performance error metric which measures the closeness of forecast distribution and corresponding observation. This score is widely used for checking the forecast accuracy. The CRPS is related to the rank probability score but compares a full distribution with the observation, where both are represented as cumulative distribution functions. Though CRPS not being used in this thesis but would be used in future research work.

GP models	WT- A					WT-B				
	MAE	MSE	MAPE	RMSE	R^2	MAE	MSE	MAPE	RMSE	R^2
No pre-correction and air density not included in the GP model	18.01	568.07	9.43	23.83	0.87	5.19	43.09	2.46	6.56	0.98
No pre-correction but with air density included within the GP model	**16.81**	**506.96**	**8.95**	**22.51**	**0.89**	**4.62**	**33.01**	**2.20**	**5.74**	**0.98**
Pre-correction applied but without air density in the GP model	18.50	598.09	9.68	24.45	0.87	4.73	35.46	2.25	0 .95	0 .98
With pre-correction and air density included with the GP	16.86	510.01	8.99	22.58	0.89	4.64	33.34	2.21	5.77	0.98

Table 6.3: Statistical measures of GP fitted models under different air density approaches

6.7.3 Residual distribution analysis using QQ plots

The distribution of the residuals can also be informative in the context of GP models. Figures 6.10 and 6.12 are the QQ plots of different air density compensation approach. The effect of different air density compensation schemes on the GP model residual distribution is significant and seen by distribution fits for all the data sets as shown in Figures 6.11 and 6.13.

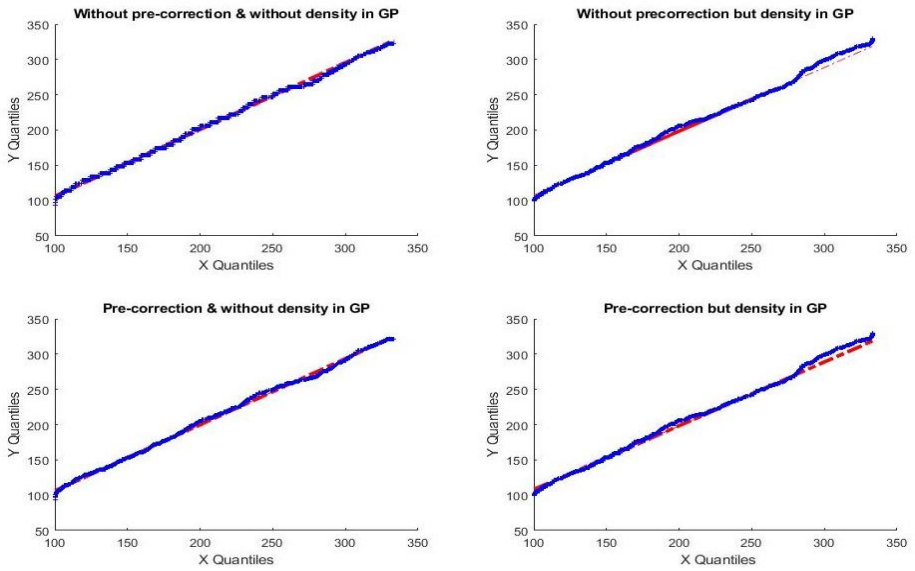


Figure 6.10: QQ plot for restricted datasets, A

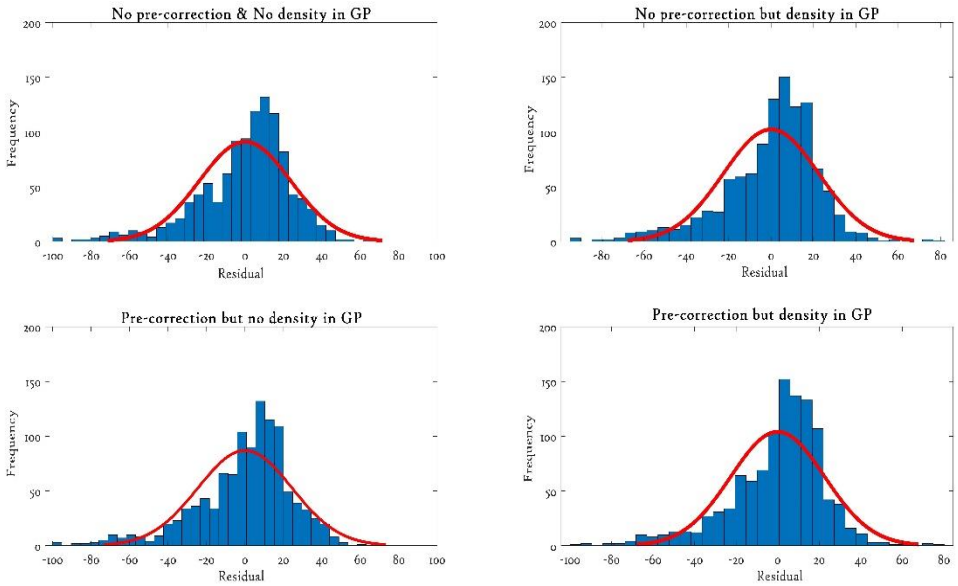


Figure 6.11: Residual histogram with distribution fit, A

Recall from chapter 5, A quantile-quantile or QQ plot is a simple graphical method used to compare collections of data with possible theoretical distributions and can help in distribution function identification. Therefore, the QQ plots are used in this chapter for GP power curve distribution analysis for different air density compensation techniques. For GP models, the residuals should be Gaussian distributed. QQ plots comparing the residual distribution with a Gaussian distribution have been calculated for the four different cases and assessed regarding RMSE differences from the ideal unity gradient line**.** Table 6.4 shows that the case of no pre-correction but with air density include within the GP model results in residuals most closely conforming to a Gaussian distribution for both wind turbines A and B. This adds to the case that the GP model is more accurate with no pre-correction, but with air density included within the GP power curve model.

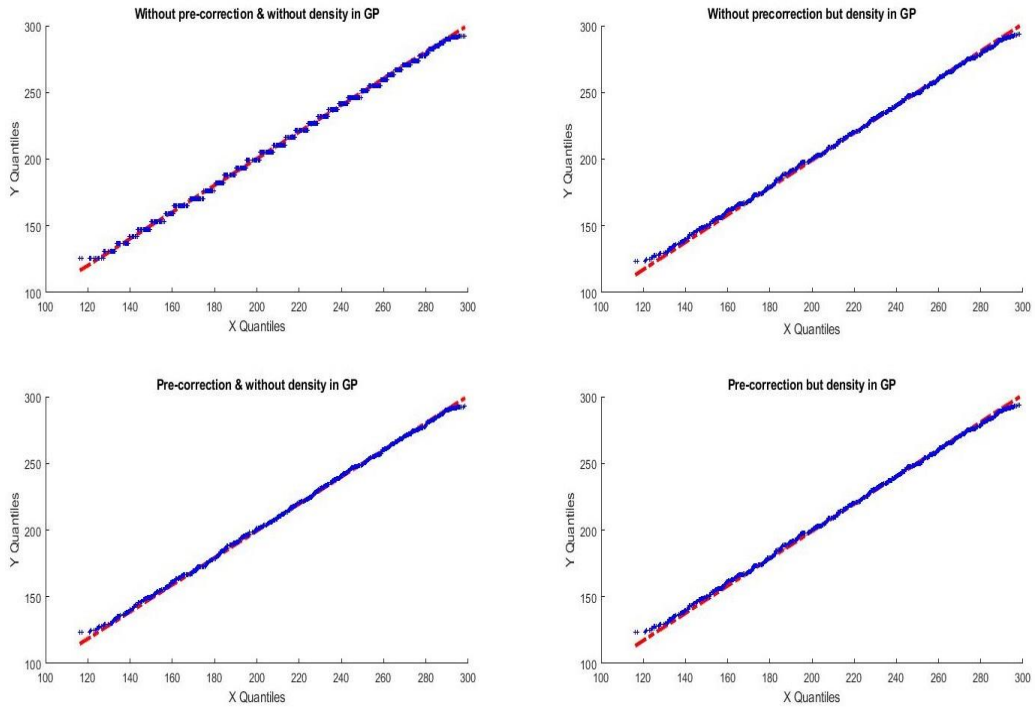


Figure 6.12: QQ plot for restricted datasets, B

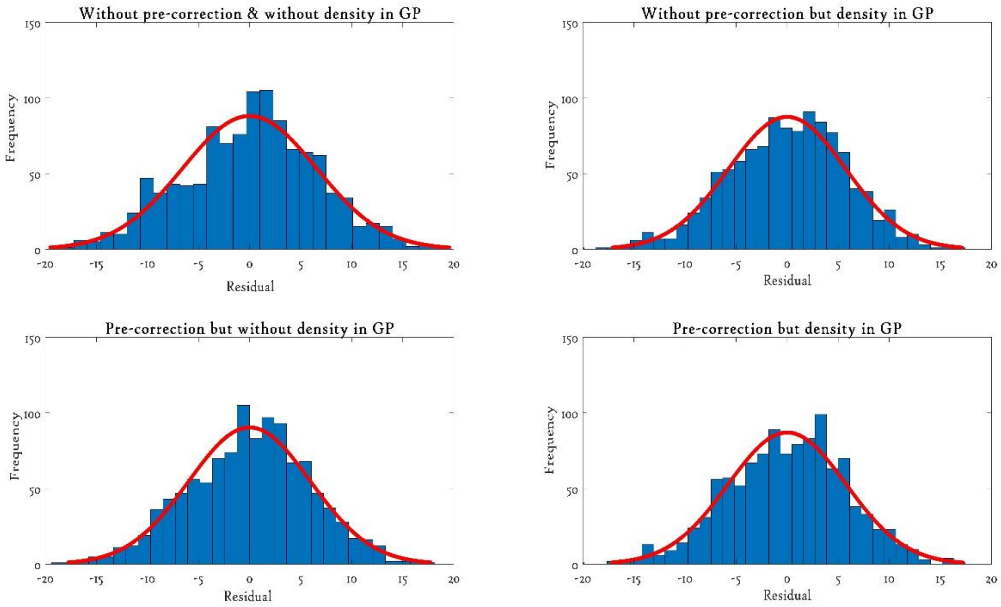


Figure 6.13: Residual histogram with distribution fit, B

Sr.no	GP models	site A	site B
1	No pre-correction and air density not included in the GP model	2.292	1.530
2	No pre-correction but with air density included within the GP model	**0.641**	**0.475**
3	Pre-correction applied but without air density in the GP model	2.020	0.545
4	With pre-correction and air density included with the GP	0.650	0.476

Table 6.4: RMSE for different density approaches from QQ plot.

6.8 Chapter conclusions

Air density corrections are essential for accurate power curve modelling. With the possibility to include air density within GP models, various alternative approaches are possible. This chapter has proposed four air density correction approaches and assessed their differing impact on GP WT power curve model accuracy. Two SCADA datasets were used that satisfied the IEC Standard guideline which states that the air density correction should only be applied when the site density differs from the standard value ($1.225\ kg/m^3$) by more than 0.05 kg/m^3. The results demonstrate that the best option is not to use the IEC Standard correction before applying the GP model. Error performance metrics suggest that separately adding density directly into the GP power curve model, without any pre-correction gives the most accurate results, see Table 6.3. Furthermore, QQ plot analysis, as provided in Table 6.4, confirms this result. In short, from this chapter we conclude that air density corrections can be significant; the best option is not to apply the Standard IEC correction before applying the GP model. This will not only help in power curve modelling but also supports the construction of robust algorithms for anomaly detection via CIs.

6.9 Chapter references

144. Jianzhou Wang, Tong Niu, Haiyan Lu, Zhenhai Guo, Wendong Yang, Pei Du. An analysis-forecast system for uncertainty modeling of wind speed: A case study of large-scale wind farms. Applied Energy. 2018; 211: 492-512. doi: 10.1016/j.apenergy.2017.11.071.

145. M. Jafarian and A. Soroudi et al. The effects of environmental parameters on wind turbine power PDF curve. In: Proceedings of 2008 Canadian Conference on Electrical and Computer Engineering, Niagara Falls; 2008, pp. 001193-001198. doi: 10.1109/CCECE.2008.4564727.

146. Abolfazl Pourrajabian, Masoud Mirzaei, Reza Ebrahimi and David Wood. Effect of air density on the performance of a small wind turbine blade: a case study in Iran Journal of Wind Engineering and Industrial Aerodynamics. 2014; 126:1-10. doi: 10.1016/j.jweia.2014.01.001.

147. Dănuţ Zahariea and Dorin Husaru. Atmospheric air density analysis with Meteo-40S wind monitoring system. In: Proceedings of 21st Innovative Manufacturing Engineering & Energy International Conference – IManE&E; 2017. Article no: 07020 .pp:6. doi: 10.1051/matecconf/201711207020.

148. G. Ciulla and V. Dio et al. Effects of the air density value on a wind generator electricity production capability. In: Proceedings of 2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC); 2016; Florence, 7-10 June, pp. 1-5. doi: 10.1109/EEEIC.2016.7555512.

149. Sungmoon Jung and Soon-Duck Kwon. Weighted error functions in artificial neural networks for improved wind energy potential estimation. Applied Energy. 2013; 111:778–90. doi: 10.1016/j.apenergy.2013.05.060.

150. Hau E. Wind turbines. Fundamentals, technologies, application, economics.2013. 3rd edition. New York: Springer publishers.
151. Chi Yan and Cristina Archer. Assessing compressibility effects on the performance of large horizontal-axis wind turbines. Applied Energy. 2018; 212:33–45. doi: 10.1016/j.apenergy.2017.12.020.
152. Liu Y.Z and Liu W.X. Effect of air density on output power of wind turbine. Applied Mechanics and Materials. 2013;(336–338):1114–1117. doi: 10.4028/www.scientific.net/AMM.336-338.1114.
153. Andrew Kusiak and Haiyang Zheng et al. Models for monitoring wind farm power. Renewable Energy. 2009; 34(3): 583-590. doi: 10.1016/j.renene.2008.05.032.
154. V. Bulaevskaya and S. Wharton et al. Wind power curve modeling in simple and complex terrain using statistical models. Journal of Renewable and Sustainable Energy.2015; 7(1): 013103. doi:10.1063/1.4904430.

Chapter 7

Anomaly detection using Gaussian Process models

In this chapter, an anomaly detection algorithm based on Gaussian Process (a non-parametric machine learning approach) models is proposed. The standard IEC binned power curve together with individual bin probability distributions can be used to identify operational anomalies. The IEC approach can also be modified to create a form of the real-time power curve. Therefore, two algorithms based on binning will be compared with a Gaussian Process anomaly detection algorithm to assess both speed and accuracy of anomaly detection. Power loss and downtime due to wind turbine rotor yaw misalignment are significant and are therefore presented as a case study to demonstrate the effectiveness of the proposed algorithms and allows the advantages and limitations of the proposed methods to be determined.

7.1 Background and Motivation

The primary role of wind turbine (WT) power curves is to provide a benchmark for performance for use in purchase contracts. The measurement and data analysis procedures have been developed over decades and are represented in an International Standard, IEC [141]. There remain issues of accuracy, and it is recognised that changes in external factors such as atmospheric stability, and associated with this wind shear and turbulence, can influence the power curves in ways that are not fully understood [155,156]. The power curves are conventionally calculated using the 'method of bins' technique prescribed by the IEC Standard and briefly described in chapter 5. This involves binning the data into 0.5 m/s wide wind speed intervals, and then calculating the average power and wind speed for each bin. The power curve is a smooth curve drawn through these points, but in actuality is only defined precisely at the points themselves. Many papers have been published that seek an alternative approach to fitting power curves to the data. These fall broadly into two categories: parametric methods such as a polynomial curve fitting; and non-parametric methods, often using machine learning techniques. These were briefly described in chapter 2.

Kernel methods are non-parametric and a supervised learning form of pattern recognition related to Gaussian Processes (GPs); they have been applied to wind turbine power curves. In [157] data from two turbines were analysed and the identification of blade icing was demonstrated. The same authors have also explored an approach based on eigenvalue analysis, [158]. A kernel plus method has also been applied, [159], to overcome identified shortcomings of a multivariate kernel regression. Air density, turbulence intensity, and wind shear are the additional variables considered. Similar data has been analyzed by the same team using the more conventional additive multivariate kernel method, [160].

Many of the published methods concentrate solely on the power curve, but apparently there is valuable information in the spread of data, for example, the probability distribution of power points in a given wind speed bin of a conventional power curve. Both Gaussian Process theory and Copula models intrinsically provide confidence bounds for the fitted power curve, and these are dealt with in more detail below.

Copula models provide an effective means of representing a non-linear relationship between two variables; they can thus be applied to wind turbine power curves. An early paper, [88], outlines how Copulas can be fitted to wind turbine power curve data. In [161], Copula fitting has been used to exclude outliers when calculating power curves.

Gaussian Processes (GPs) are very general non-linear models where training is undertaken to select the best hyperparameters which define the covariance function. Comparative studies of stationary covariance functions have been undertaken in chapter 3. The results show that rational quadratic functions perform almost as well as squared exponential covariance functions although prediction speed and training time took are longer. Therefore, the squared exponential covariance function based GP model gives the relatively better results and is used for all GP models from now onwards.

This chapter proposes a range of algorithms for condition monitoring based on probabilistic GP and binning methods. A comparative assessment of promising approaches has been undertaken in terms of their capability to detect in advance (and by how much) signs of failure.

7.2 Chapter novel contributions

In this chapter, we propose a novel approach based on a Gaussian Process (GP) for anomaly detection. Yaw misalignment is used as a case study for this work. To assess the GP model effectiveness, two other methods, based on binning are constructed and then compared with the GP fault detection algorithm. The outcomes of this chapter demonstrate that the GP model was able to detect the anomaly effectively with the alarm raised only 1.5 hrs after the fault first occurred. Not only is the GP method able to detect the yaw misalignment quickly, but it produces no false positives, in contrast to other bin based models (described in upcoming sections), hence confirming that the GP approach provides both fast and robust fault identification. The highlights of this chapter are as follows:

- A SCADA based Gaussian Process model for anomaly detection of a wind turbine.
- Two probabilistic binning methods developed for validating the effectiveness of the GP model.
- A case study exhibiting wind turbine rotor yaw misalignment is presented and used for testing the proposed models.
- Gaussian Process models are shown to be able to detect anomalies quickly and without any false positives.

7.3 SCADA datasets description

The SCADA data sets used in this study are from operational wind turbines located in Scotland, UK, and contain more than 100 different signals, ranging from the timestamp, calculated values, set points, measurements of temperature, current, voltage, wind speed, power output, wind direction and so on.

Due to sensor failure and data collection faults SCADA data is not itself without some errors. Such errors will affect power curves and should be systematically removed at the outset. Criteria briefly described in section 2.5 such as timestamp mismatches, out of range values, negative power values, and turbine power curtailment are considered to be misleading and have been filtered out as summarised in Table 7.1. All the SCADA data considered here consists of 10-minute averages with maximum, minimum, and standard deviation over the 10 minutes also being recorded. The data used in this chapter is coming from 2.3

Data set	Start timestamp	End timestamp	Description
1	11/1/2008 14:30 PM	30/03/2008 15:20 PM	Total data filtered set: 3274 observations
2	14/4/2009 11:20 AM	16/4/2009 9:50 AM	Total data set: 201 observations

Table 7.1: Description of the SCADA datasets

MW Siemens turbines. Table 7.1 summarises two datasets: dataset 1 beginning with time stamp ‘‘11/1/2008 14:30 PM’’ and ending at time stamp ‘‘30/03/2008 15:20 PM’’; and data set 2 beginning with time stamp ‘‘14/4/2009 11:20 AM’’ and ending at time stamp ‘‘16/4/2009 9:50 PM’’. Dataset 1 contains 4725 measured values which reduced to 3274 data points after pre-processing and was used to develop power curve models based on binning and a GP, as described in upcoming sections. Dataset 2 included 201 data points and was used to test the performance of the model learned from dataset 1.

7.4 Yaw misalignment a case study

Wind direction is not constant; it always changes. A wind turbine generates most effectively when the rotor is facing directly into the free oncoming wind; the yaw drive and control system tracks changes in wind direction. A wind vane, located

on the nacelle behind the rotor blades (Figure 7.1), is used to control the alignment of the wind turbine. Nevertheless, an erroneous yaw error signal can result from turbulence generated with the passage of the blades despite sophisticated corrective software, and this will result in a degree of so-called yaw error where the turbine rotor is not correctly aligned. The wind vane measures the yaw error mounted on the rear of the nacelle. Significant yaw misalignment, reflecting a yaw control error or fault, results in a loss of power, and additional rotor loads. Such a fault is quite common, and early detection is essential to prevent loss of power generation and excessive fatigue damage.

In addition to the challenge of measuring wind direction on the nacelle, any misalignment of wind vane itself will contribute to yaw error. Moreover, large wind turbine rotors must be yawed slowly to limit gyroscopic loads and thus cannot follow rapid changes in wind direction. All this means that some level of operational yaw error is inevitable. This average yaw misalignment will result in some reduction to the annual energy production (AEP). As reported in [162] an average of 6.2° of misalignment causes an estimated 2% reduction in AEP with roughly a loss of 2 to 3 % for 9° to 10° of average yaw error. As well as the operational yaw error outlined above, faults can develop with the yaw control system. Early detection of such yaw control faults is vital to avoid loss of power production and associated revenue [163], but also to minimise fatigue damage and reduce maintenance costs, [164], increase life of turbine, [165], and reduce the levelized Cost of Electricity (LCOE) and improve the return on investment (ROI), [166].

The dependence of wind turbine power production on yaw error is reasonably explained by the cosine cubed theory, [168], which states that power output is scaled by the cube of the cosine of yaw error (differences between wind direction and the nacelle direction). Although not perfect, this law estimates that a large yaw error of 20° will lead to a significant power deficit of 17%. Such power

deficit is unlikely to acceptable to a wind farm operator. Because of the strong link between yaw error and power production, such faults can provide excellent test cases with which to test power curve anomaly detection algorithms.

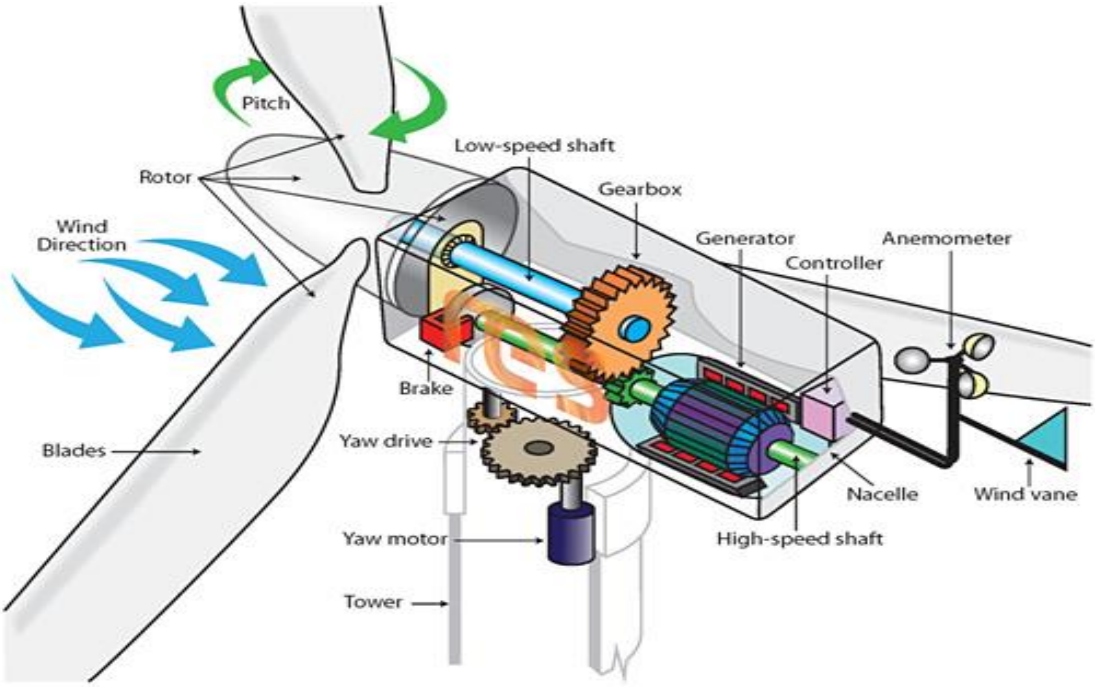


Figure 7.1: Schematic of wind turbine system, [167]

7.5 Construction of reference power curve using Gaussian Process

In recent years, the importance of power curves in the potential use of wind turbine condition monitoring has been recognised and very attractive from the industry point of view. Significant changes in the power curve can be an indication of performance change most likely due to faults. SCADA based power curve algorithms are cost-effective and widely used for the identification of performance and or condition changes. As part of this, chapter 5 described the different new methods based on power curve and how these can be used for analysing the turbine performance, failure detection and power optimisations briefly. In most cases, the SCADA based power curve data is 10 minutes averaged pairs of net wind turbine power, and hub height horizontal wind speed

at a suitable distance in front of the rotor (between 2 and 4 rotor diameters in the Standard), or, alternatively taken from the nacelle anemometer.

10-minute SCADA data from a healthy turbine are utilised for reference power curve construction. These air density corrections (outlined in section 5.4) also applied to the unhealthy turbine data so that it can be evaluated against the reference power curve in a consistent manner. The power curve data shown in Figure 7.2 is after the filtration and air density correction. Using the GP theory (described in chapter 3), a power curve is constructed from filtered, and air density corrected SCADA data and is shown in Figure 7.3. As can be seen clearly in Figures 7.2 and 7.3, the GP power curve is smooth (smoothness rough estimate is done by calculating the standard deviation of the differences see figure 6.7), accurate and represent the expected characteristics of the power curve. It should be noted that the estimated power curve is based on held out data.

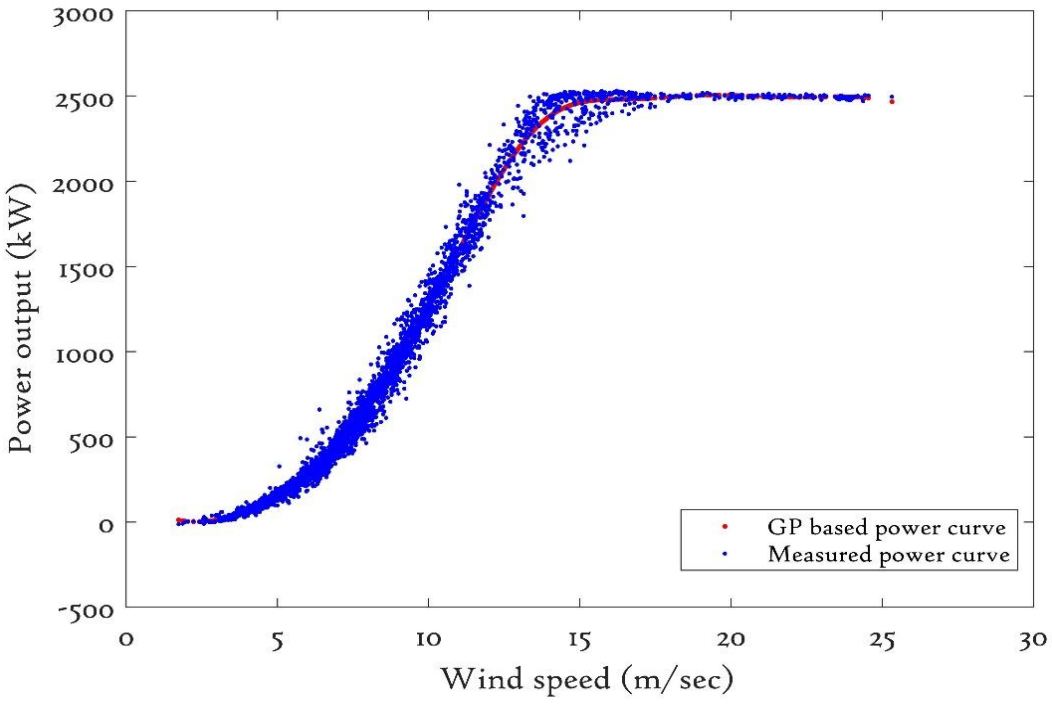


Figure 7.2: GP power curve comparison

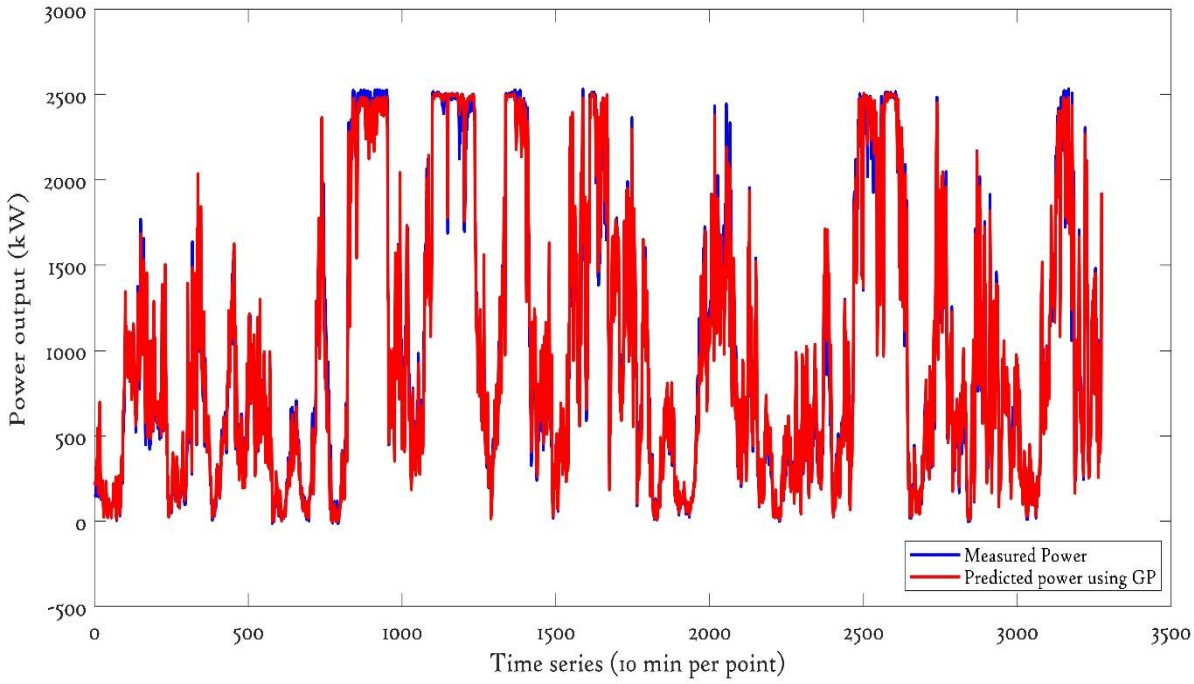


Figure 7.3: GP estimates of power generation time series

7.6 Methodologies to be compared

Wind turbine performance analysis is presently an active area of research. This chapter is focused on performance as represented by fitted power curves.

- **Power curve comparison**

Power curves can play a vital role in identifying anomalous operation reflecting insipient fault development. Information related to statistically significant power curve deviations is thus essential. Such deviations could be due to anemometer error, power transducer calibration error or controller setting error. A consistent significant deviation from the reference power curve may be due to blade damage; yaw drive issues or possibly substantial wind shear or wind veer for below-rated operation, [169]. The IEC recommended approach uses binning; hence in order to evaluate the effectiveness of the GP model, two binning methods are described and a comparative analysis carried out. It is worth noting that the power deficit caused by yaw error is unlikely to be acceptable to a wind farm operator; hence, a yaw error threshold of 20 degrees has been chosen for binning algorithms.

Here, the reference power curve based GP and binning algorithms were constructed for failure detection. Identifying any statistically significant deviation from the reference power curve on the bin by bin basis for the binned reference power curve, and for the GP model using the inherent confidence intervals (described in chapter 3) approach is being adopted. The term 'incoming' is being used in the upcoming section that refers to the unhealthy SCADA data that tested against developed binning and GP model.

- **Fisher's combined probability test**

Statistical tools often need to combine the evidence obtained from assumed independent sources. In order to do so, a combined p-values concept is developed. The Fisher product test, [170], is a statistical test that combines p-values, based on the notion that several non-significant results occurring together may suggest significance and hence detect departure from the null hypothesis, H_o. The equation used for calculating the Fisher combined probability test is given as,

$$X_{2k}^2 \sim -2\sum_{i=1}^{k} ln\, p_i \qquad (7.1)$$

Where p_i is the probability that the i^{th} variable exceeds the measured value under the null hypothesis. Under the null hypothesis X_{2k}^2 is the distributed as a chi-squared variate with $2k$ degrees of freedom. Here, k is the number of independent tests being performed.

Apart from the Fisher method, other methods for combining p-values are briefly discussed in [171,172,173]. The applications for combining the p-values are numerous, including combining the results from independent studies and combining the results of individual component problems as part of an overall test, [174,175]. This approach will be used to construct an effective probabilistic based binning and for the GP algorithm, as is described in the upcoming sections.

For consistency, all the time series used in this work covers the same time period of 33 hours. On the graphs these are shown on a scale of zero to 200 (10-minute periods); in reality, the time series runs from 11:20 on 14-04-2009.

7.6.1 Probabilistic assessment of incoming data using a binned power curve

This algorithm uses a probabilistic approach to identify anomalies associated with incoming power-wind speed data, point by point. A reference binned power curve has been constructed following the IEC recommendations using the pre-processed power curve (of Figure 7.2) where it is binned by the standard 0.5 m/sec wide wind speed intervals, see chapter 5. Figure 7.4 shows the reference binned power curve together with error bars. The two standard deviations (i.e., 95% confidence intervals) of measured power values are used to calculate the error bars which are used to measure the uncertainty associated with each bin of the power curve. It is worth noting that there is a slight deviation in power curve which is due to varying environmental conditions.

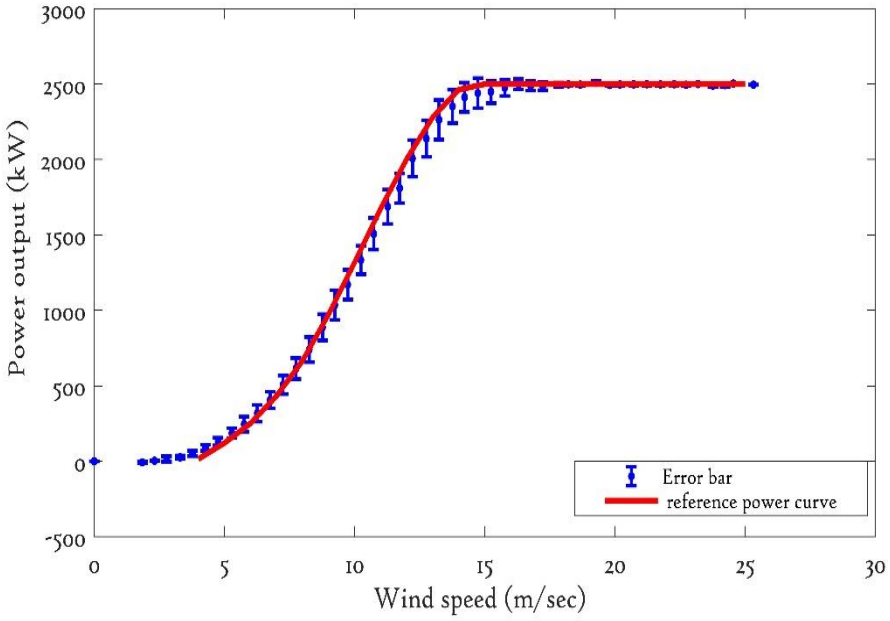


Figure 7.4: Binned power curve with error bars

Once the reference power curve has been constructed, the next step is to assess potentially unhealthy incoming data point by point against the appropriate bin and its uncertainty by probabilistic assessment. In the probabilistic binning

method, Fisher's combined probability test (described above) is used to multiply two sequential p values. After testing and validating various different numbers of p-values, it was found that two p-values gives the most effective performance, and hence will be used here. Two sequential values have been used here, but this could be extended if required to give greater confidence that the data points are anomalous, but that will affect the model performance in identifying the false alarm shown in this case. A threshold of 0.005 (or significance level) is fixed in order to determine the aggregate number of anomalous bin values.

A graph of yaw error together with alarm detection based on the probabilistic approach is plotted in Figure 7.5. From the graph, an anomalous performance due to yaw misalignment or a yaw drive control issue can be identified. This is further confirmed in Figure 7.6, where it can be seen that the nacelle is stuck in a fixed position for an extended period of time. It can be observed that the yaw error exceeds 20° consistently for timestamps 50 to 100. This anomalous performance was confirmed to be a yaw drive or yaw drive control issue.

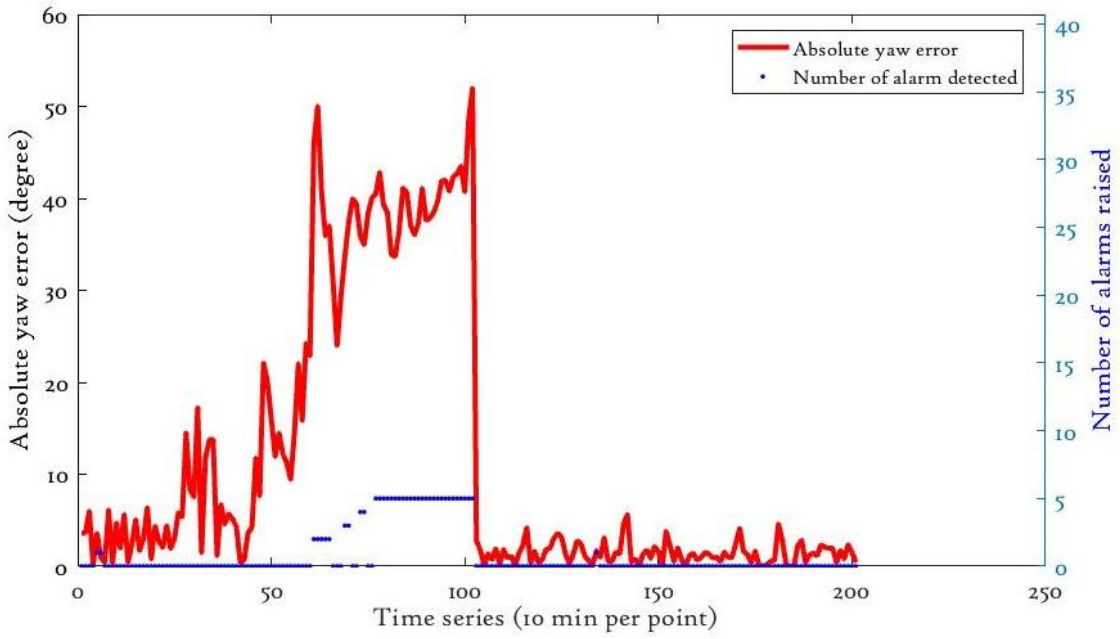


Figure 7.5: Yaw error detection using a probabilistic assessment of binned power curve

Using this method, the first alarm recorded at 00:50 15-04-2009, approximately 4 hours after the yaw fault was first identified (in Figure 7.5) at 21:00 14-04-2009. It should be noted that this time series also indicates a limited number of false alarms. Overall though, the proposed approach, based on the probabilistic binned power curve approach, allows identification of this significant performance issue and does this relatively quickly.

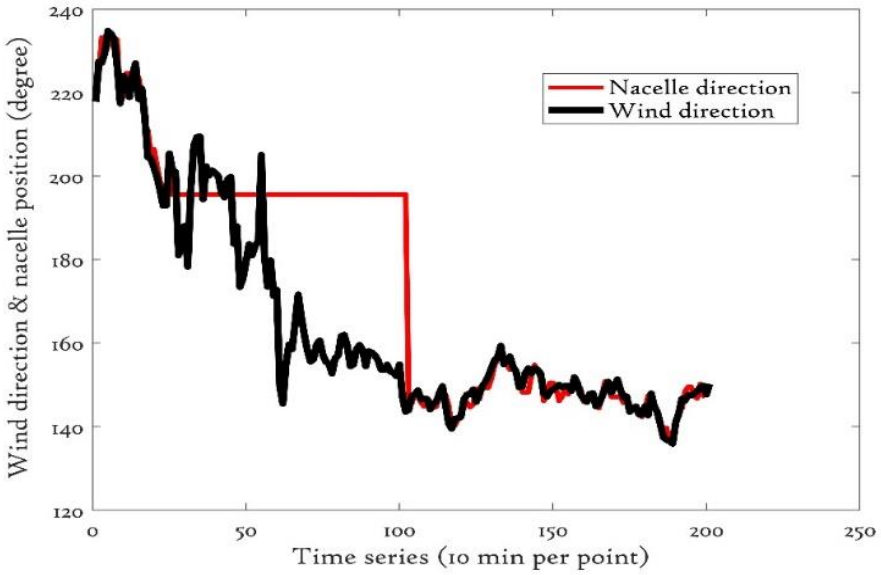


Figure 7.6: Time series of wind direction and nacelle position

7.6.2 Probabilistic assessment of incoming data using a real-time power curve

Wang and Infield, [176], proposed an approach to anomaly detection by comparing a so-called real-time power curve with a reference power curve using Welch's hypothesis test with a confidence interval of 99.5 %, i.e., a significance level of 0.005. Selection of this provisional figure was on an intuitive basis where the calculated likelihood is not unreasonable. A significance level of 0.005 was used to determine the aggregate number of anomalous bin values at each time point for an entire power curve. These are plotted as a time series as shown in Figure 7.7 and record the anomalous performance due to the yaw drive or yaw

drive control fault. Using this algorithm, an alarm would have been raised at 03:00 on 15/04/2009, 6 hours after the fault occurs (at 21:00 on 14/04/2009).

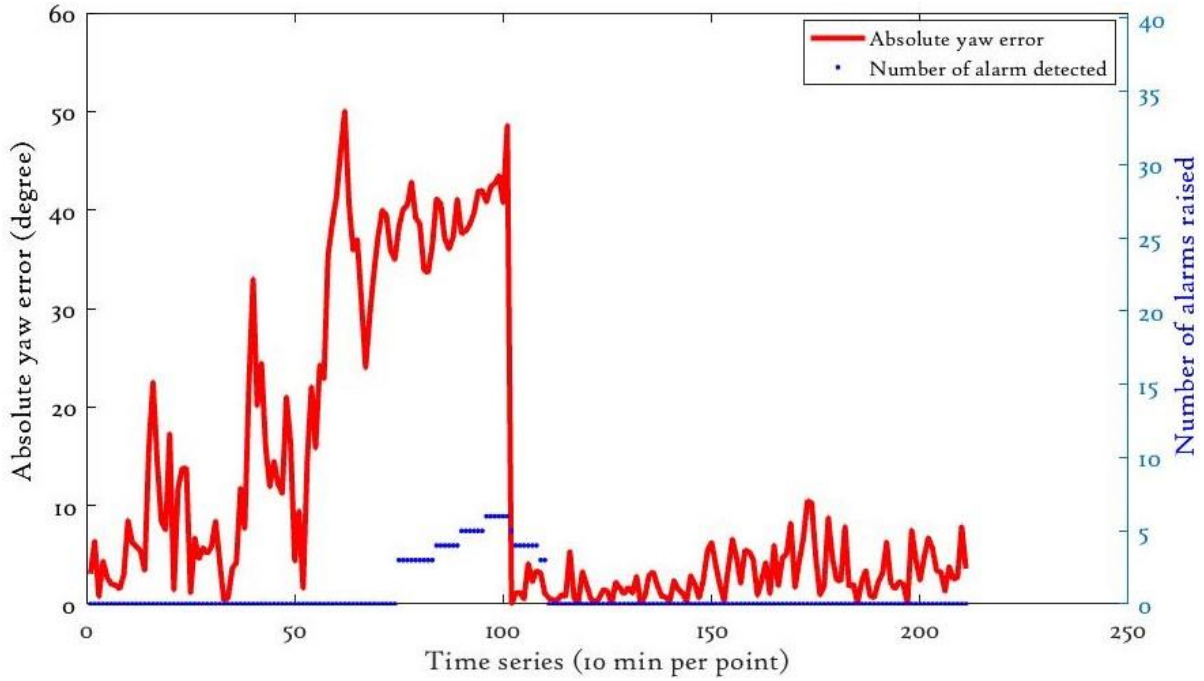


Figure 7.7: Absolute yaw error detection using online power curve model

7.6.3 Probabilistic assessment of incoming data using a Gaussian Process to represent the power curve

A GP power curve model intrinsically represents fitting errors and thus model accuracy where confidence intervals (CIs) are used to assess the model performance; the calculation and importance of CIs with regards to GP models uncertainty are briefly described in chapter 3. Here, CIs are used to assess the GP power curve, and the intervals provide information on the uncertainty surrounding an estimation but are themselves model-based estimates. Data points that lie outside of the confidence intervals can be considered anomalous, signifying a potential malfunction of the wind turbine. The GP power curve with CIs is shown in Figure 7.8, and this will be used as the reference GP power curve to detect the anomaly caused by yaw error misalignments.

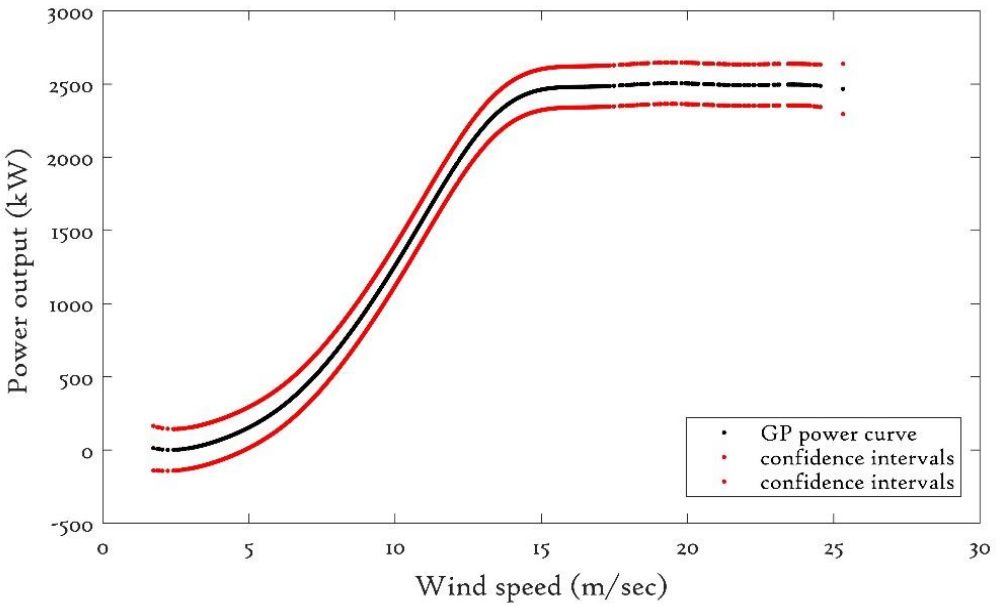


Figure 7.8: Reference power curve fitting with confidence intervals using GP

The CIs of 95 % reflects a significance level of 0.05 and was used in the GP model to determine the sequential anomalous data point values at each time for an entire reference GP power curve. The incoming data points are assessed on a point by point basis against the reference GP power curve and probabilistic assessment undertaken. The Fisher test as described above was used to combine 2 p-values with a threshold of 0.008 applied to filter the individual p-values. False alarms cause additional operational costs, and so it is desirable to construct an effective GP model that generates no false alarms. To achieve this, the GP is model adjusted by changing the probability threshold until no false alarms occur; a threshold of 0.008 gives accurate results with no false positives but slightly affect the GP model capability to detect the anomaly but does not affect the GP model overall performance and hence this threshold is used here.

Figure 7.9 shows the effectiveness of the GP algorithm where the aggregate number of anomalous values at each time point is plotted as a time series together with absolute yaw error. Using the GP algorithm, an alarm would have been raised at 22:40 on 14/04/2009, just 1.5 hrs after the start of the yaw fault at 21:00 on 14/04/2009, as shown in Figure 7.9. Moreover, as described above, there are

no false positives, confirming that the GP approach provides both fast and robust fault identification.

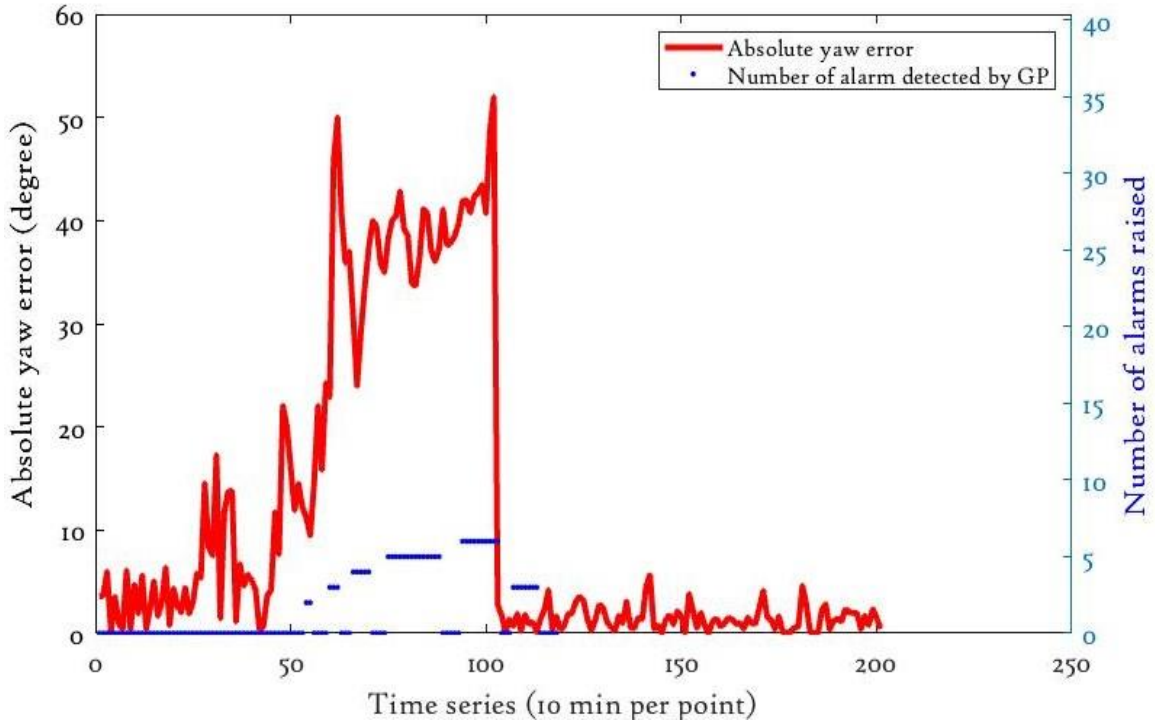


Figure 7.9: Absolute yaw error detection using GP model

7.7 Performance comparisons of different methodologies

A performance comparison of the three different approaches to detect power performance anomalies associated with a yaw control fault, in terms of speed of detection, are presented in this section. It is essential to detect the yaw error as quickly as possible otherwise if it remains undetected for a long time then it yields large yaw error that ultimately leads to enhanced damaged rates; hence the speed of detection here is given importance for detection performance comparison purposes.

The response of the three models are brought together and plotted against the yaw error time series as shown in Figure 7.10. As can be identified from the nacelle position time series (Figure 7.6), the yawing fault started at 21:00 on 14/04/2009 and hence this will be used as the benchmark to judge the capability each of the models' capability to rapidly detect the fault.

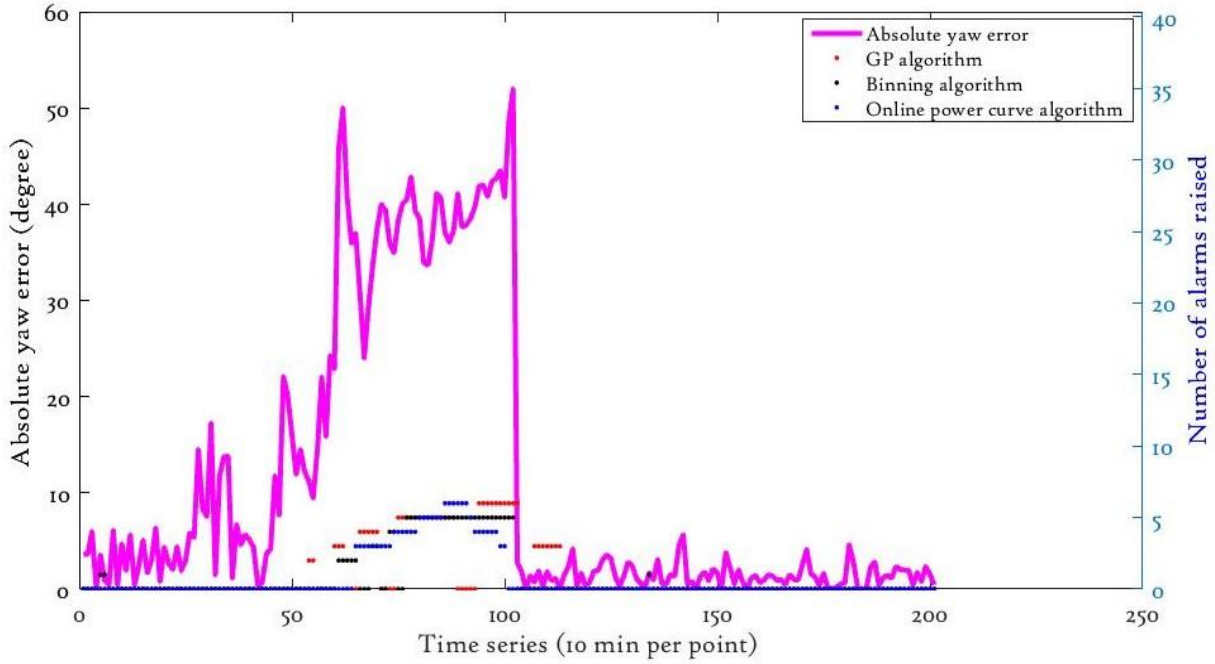


Figure 7.10: Comparative analysis of different models for absolute yaw error detection

By doing comparative analysis, it has been found that the GP approach is able to detect yaw error earlier, just 1.5 hours after the fault occurs, while probabilistic binning approach took around 4 hours, and the real-time power curve algorithm took 6 hours to identify the fault, as shown in Figure 7.10 and Table 7.2. Not only is the GP method able to detect the yaw misalignment quickly, but it produces no false positives, in contrast to the methods using the probabilistic binned power curve.

Models	**Alarm detected**	**Time taken to identify the fault**
Online power curve model	3:00 on 15/4/2009	6 hours
Probabilistic assessment using binning	00:50 on 15/4/2009	~ 4 hours
Probabilistic assessment using GP	22:40 on 14/4/2009	1.5 hours

Table 7.2: Alarm record and detection by each approach

7.8 Chapter conclusions

GP models have been shown to provide an effective approach to power curve based condition monitoring. Models based on GP and binning have been analysed to assess their effectiveness in terms of capability to detect in advance (and by how much) anomalous performance related to yaw misalignment. Accurate modelling of the reference power curve is vital in the development of algorithms to detect such anomalies. The result of this chapter demonstrated that the GP fault detection algorithm detects anomalies faster, i.e., at 22:40 on 14/04/2009, just 1.5 hours after the fault, which is considerably quicker than the two binning algorithms (took 4hrs and 6 hrs respectively).

From turbine operator's point of view, it is therefore vital to increase the effort spent on monitoring turbine health to reduce unscheduled downtime and operational costs. This early detection is helpful to avoid early replacement of healthy components and reduces the risk of reaching the catastrophic stage that causes significant downtime and high operational cost. This early detection also gives enough time for maintenance action in advance, as well as optimise the performance of wind turbines. The GP fault detection algorithm does not produce false positives, unlike the binning algorithms. Also, binning algorithms compromise accuracy due to the bin width selection because within each bin the measured power will depend strongly and non-linearly on wind speed and a wide bin would result in a systematic bias, and the need in practice to get sufficient data points in each bin to be of statistical significance.

While dealing with GPs, it is important to note the data management challenge. Due to the inverse cubic problem, it is not desirable to include a large number of SCADA data points in the models, either for training or fitting. Some methods can help with this inverse cubic problem, but these still require high processing power and computational cost in dealing with large SCADA datasets. Hence striking a balance between these two is vital for GP model accuracy.

7.9 Chapter references

155. St. Martin, C. M., Lundquist, J. K., Clifton, A., Poulos, G. S., and Schreck, S. J.: Wind turbine power production and annual energy production depend on atmospheric stability and turbulence, Wind Energ. Sci., 1, 221-236. doi: 10.5194/wes-1-221-2016. 2016.

156. Emil Hedevang. Wind turbine power curves incorporating turbulence intensity. Wind Energy 2014; 17:173–195.

157. Georgios Alexandros Skrimpas, Christian Walsted Sweeney, Kun S. Marhadi, Bogi Bech Jensen, Nenad Mijatovic and Joachim Holbøll. Employment of Kernel Methods on Wind Turbine Power Performance Assessment. IEEE Trans. Sust. Energy, Vol. 6, No. 3, July 2015, pp 698-706

158. Georgios Alexandros Skrimpas, Christian Walsted Sweeney, Kun S. Marhadi, Bogi Bech Jensen, Nenad Mijatovic, and Joachim Holbøll. Detection of Wind Turbine Power Performance Abnormalities Using Eigenvalue Analysis. Annual Conference of the Prognostics and Health Management Society, 2014

159. Giwhyun Lee, Yu Ding, Le Xie and Marc G. Genton. A kernel plus method for quantifying wind turbine performance upgrades. Wind Energy 2015; 18:1207–1219.

160. Giwhyun Lee, Yu Ding, Marc G. Genton, and Le Xie. Power Curve Estimation with Multivariate Environmental Factors for Inland and Offshore Wind Farms. Journal of the American Statistical Association, 110:509, March 2015, pp 56-67

161. Wang, Y., Infield, D., Stephen, B. & Galloway, S. Copula based model for wind turbine power curve outlier rejection; Wind Energy, Vol.17, Issue 11, pp 1677-1688, Nov. doi: 10.1002/we.1661. 2014.

162. 'Flexible solutions to optimize turbine performance'. (accessed 15th July 2018).

163. ‘Optimising annual energy production with apt handling of yaw misalignment’ 4th Wind Operations and Maintenance Canada 2017. 4th November,2013.
164. M.T. van Dijk, J.W. van Wingerden, T. Ashuri, Y. Li, M.A. Rotea. Yaw-misalignment and its impact on wind turbine loads and wind farm power output. J Phys Conf Ser, 753 (6) (2016), p. 062013.
165. ‘Wind turbines: correcting yaw error’. (accessed 19th August 2018).
166. ‘How to improve the LCOE of wind power’. (accessed 19th August 2018).
167. ‘Renewable energy sources: wind’. (accessed 19th August 2018).
168. Mamidipudi, P., Dakin, E., Hopkins, A. F., Belen, C., Leishman, J. G. Yaw Control: The Forgotten Controls Problem. EWEA Brussels, 2011.
169. Wagner, R. Antoniou, I. Pedersen, S. M. Courtney, M. S. and Jørgensen, H. E. (2009). The Influence of the Wind Speed Profile on Wind Turbine Performance Measurements. Wind Energy, Vol. 12, Issue 4, pp. 348-362.
170. Fisher, R.A., 1932. Statistical Methods for Research Workers. Oliver and Boyd, London.
171. Ori Davidov. Combining p-values using order-based methods. Computational Statistics & Data Analysis, Volume 55, Issue 7, 2011, pp. 2433-2444.
172. Brown, Morton B. 400: A Method for Combining Non-Independent, One-Sided Tests of Significance. Biometrics, vol. 31, no. 4, 1975, pp. 987–992. JSTOR, JSTOR. doi: http://www.jstor.org/stable/2529826.
173. James T Kost, Michael P McDermott. Combining dependent P-values. Statistics & Probability Letters Volume 60, Issue 2, 15 November 2002, Pages 183-190.
174. McDermott, M.P., 1999. Generalized orthogonal contrast tests for homogeneity of ordered means. Canad. J. Statist. 27,457–470.

175. Cox, C.P., Han, C., 1982. Testing multivariate means when the covariance matrix has intraclass correlation structure. J. Statist. Comput. Simulation 16, 97–107.

176. Yue Wang, David G. Infield. Power Curve Based Online Condition Monitoring for Wind Turbines. COMDEM 2013 conference. doi: 10.13140/2.1.4492.9928.

\

Chapter 8

Gaussian Process models incorporating additional wind turbine parameters

The IEC Standard 61400-12-1 recommends a reliable and repeatable methodology for computing the turbine performance where a data reduction approach called the 'method of bins' is used for accurate computation of the wind turbine power curve in which only mean wind speed at hub height and the air density are recognised as relevant input parameters. However, several literature studies have suggested that power production from a wind turbine depends on several critical variables, namely: rotor speed and blade pitch angle, commonly known as performance parameters of a wind turbine. Furthermore, the cubic relationship of power output and wind speed shows that the power generation of the wind turbines affected by rotor speed, blade pitch angle and therefore impact of these performance variables on GP models needs to be investigated.

This chapter presents a technique that is computationally tractable and able to demonstrate the impact of these additional wind turbine performance parameters (blade pitch angle and rotor speed) on the GP power curve model accuracy and uncertainty. Based on this analysis, an extended GP fault detection algorithm is constructed (in which yaw misalignment is used as a case study) and compared against the GP algorithm of chapter 7 in terms of their capability to detect in advance (and by how much) signs of failure, and also their false positive rate, by making use of extensive SCADA data and turbine fault and repair logs.

8.1 Motivation and chapter contributions

The International Standard IEC 61400-12-1 provides guidelines for calculating power curve using the 'method of bins', but it does not include the dynamic effects of the different variables involved in the process. The details of the binning method are described in chapter 5. As already described in chapter 4, many of the published nonparametric methods used the power curve for condition monitoring in which mean wind speed at hub height and the air density wind speed considered as suitable input parameters. However, the power output of a wind turbine is affected by various additional performance parameters, in particular, rotor speed and blade pitch angle, and also atmospheric parameters such as turbulence intensity and wind shear. Analysis of these parameters can be useful for robust condition monitoring of a wind turbine.

The power output of the turbine is directly proportional to swept area, power coefficient and wind speed where the power coefficient is a nonlinear function of the tip speed ratio (TSR) and blade pitch angle. TSR relies upon wind speed, and rotor speed, subsequently pitch angle and rotor speed affect the power performance of wind turbines. The pitch angle and rotor speed are referred to as WT performance parameters since they affect the power production of turbines. Discussion of these parameters was briefly presented in chapter 5. The atmospheric parameters (wind shear and turbulence intensity) also affect the GP models accuracy and uncertainty, but these are not included in this chapter.

The uncertainty and stochastic nature of wind power make operation and planning a challenging task; henceforth the influence of these performance variables on nonparametric models is essential in order to improve model accuracy and reduce uncertainty to construct robust fault detection algorithms that have a capability to detect failures quickly without any false positives. However, the influences of

these additional performance parameters on nonparametric models have never been explored. The research present in this chapter tries to full fill this gap.

In this chapter, GP power curve models were constructed with the inclusion of these additional wind turbine performance parameters directly from data instead of using the underlying physics and detailed analysis has been carried out. Furthermore, comparative studies are presented in order to find out the most influencing performance parameter that affects the GP model accuracy and uncertainty significantly. After finding the most prominent performance parameter, the fault detection algorithm is constructed in which the most prominent performance parameter is used. After that, the constructed algorithm is compared with the algorithm of chapter 7 to assess its effectiveness in terms of speed of fault detection and false positive alarms.

8.2 Inclusion of wind turbine performance parameters into Gaussian Process models

As already described in chapter 3, a GP model is mathematically expressed by its mean and covariance function (or kernel). The general covariance matrix, K, gives the variance of each variable along the leading diagonal, and the off-diagonal elements measure the correlations between the different variables mathematically described as follows,

$$K = \begin{bmatrix} k_{11} & \cdots & k_{1n} \\ \vdots & \ddots & \vdots \\ k_{n1} & \cdots & k_{n} \end{bmatrix} \quad \text{Where } k_{ij} = k(x_i, x_j)$$

K is of size $n \times n$, where n is the number of input parameters considered, and it must be symmetric and positive semidefinite i.e. $K_{ij} = K_{ji}$.

Due to multivariate nature of a GP, $n =$ number of predictors selected for GP model where x represent the wind speed along with wind turbine performance parameters (rotor speed and rotor speed) to facilitate analysis of the effect of performance parameters on GP models.

Using GP theory (described in chapter 3), a power curve based on the GP model incorporating the additional performance parameters (rotor speed and pitch angle), included individually in the GP model along with wind speed, is constructed. The squared exponential covariance function is used to construct GP models since it has been found to be most suitable covariance function for wind turbine condition monitoring purposes, see chapter 3. Figure 8.1 shows the estimated power and measured power values in time series where the impact of these performance parameters on GP models are clearly shown. The inclusion of rotor speed makes a significant improvement in prediction thought the inclusion of blade pitch also improves the model but not to the same extent, as shown in Figure 8.1. This leads to the investigation of performance parameter impacts on GP model accuracy and uncertainty, as described in upcoming sections.

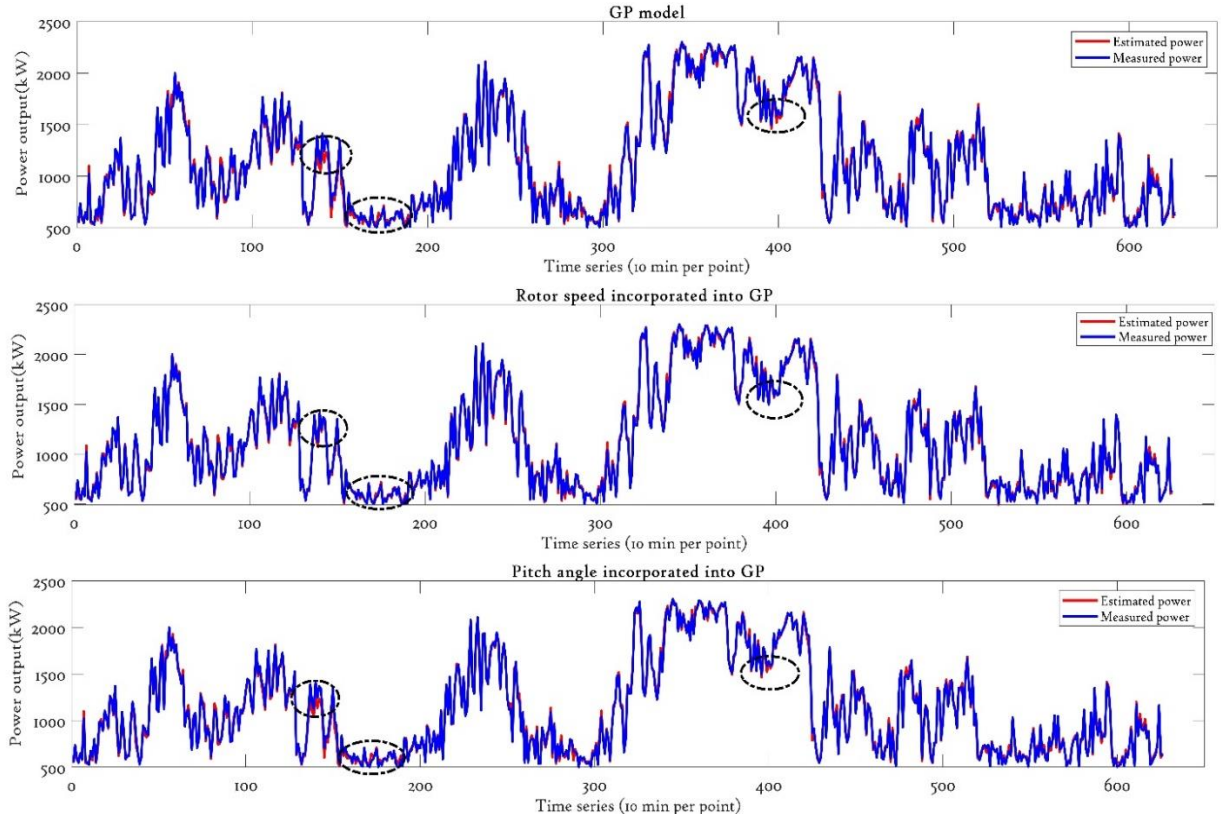


Figure 8.1: Estimated GP power curve comparison in time series with the inclusion of performance parameters

As already described in previous chapters, GP models provide a probabilistic framework for performing inference over functions. For a given observed data and new input, estimation corresponds to computing the (Gaussian) predictive distribution of the associated output, whose mean can be used as a point estimate. The obtained predictive variance gives confidence intervals (CIs) on this estimate which is useful in analysing the uncertainty associated with GP models and used here to investigate the impact of performance parameters on GP models uncertainty. It is worth noting that GP model uncertainty uses probabilistic descriptions of the model input which can be used to derive probability distributions of model outputs and system performance indices as briefly described in chapter 3.

- **Gaussian Process model incorporating blade pitch angle**

The rotor power coefficient is a function of blade pitch angle which is used to regulate power production from a wind turbine. Averaging the angles of the three blades (in case of a three-bladed wind turbine) gives the value of pitch angle, and this is adjusted under normal operation to capture the maximum power below rated power and to limit power output above rated wind speed. Therefore it affects the power production and shape of the power curve as so should be considered within the GP power curve model.

The GP based power curve obtained by theory (described in chapter 5) in which blade pitch angle is incorporated into the squared exponential covariance function along with wind speed as described in section 8.3 is shown in Figure 8.2. Figure 8.2 suggests that the inclusion of pitch angle improves the GP model accuracy and reduces uncertainty up until rated wind speed and after that its accuracy and uncertainty deteriorates; this is because above-rated wind speed blade pitch is controlled to regulate the power.

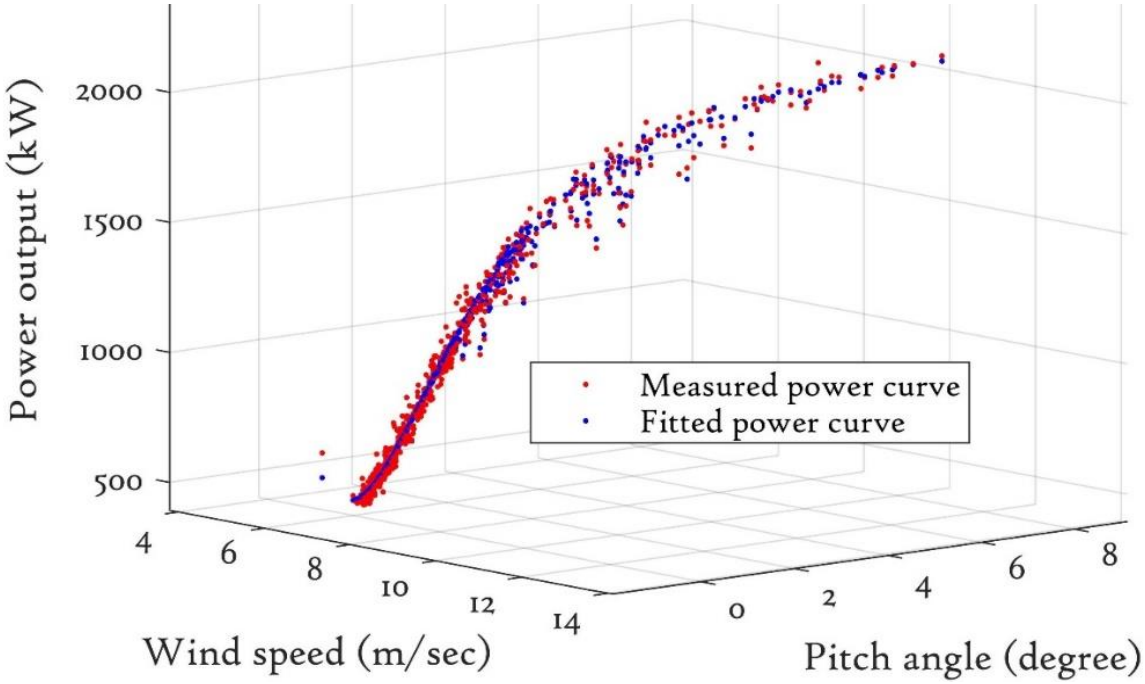


Figure 8.2: Estimated GP power curve with the inclusion of pitch angle

As already described, the calculated CIs are used for GP model uncertainty analysis; this is plotted against wind speed and compared against with GP without the inclusion of pitch angle model. Figure 8.3 suggests that when blade pitch angle incorporated GP power curve model, the model uncertainty is reduced. The improvement is seen between the cut in and rated wind speed where maximum power tracking takes place which is the most critical wind speed region for condition monitoring purposes. At above-rated wind speed, uncertainty deteriorates, but it is not an issue since above-rated speed, turbine operation control by operators.

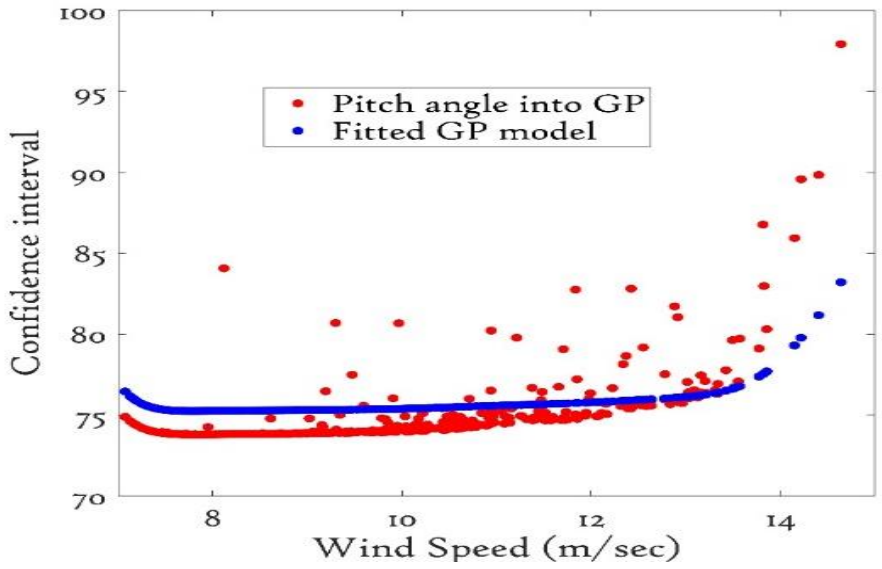


Figure 8.3: GP models Uncertainty analysis when pitch angle incorporated

- **Gaussian Process model incorporating rotor speed**

Like pitch angle, the rotor speed can be incorporated into squared exponential covariance function as per methodology described above. The resulting estimated power curve is then compared with a measured power curve as shown in Figure 8.4. The inclusion of rotor speed improves the GP power curve model accuracy across the entire wind speed range, see Figure 8.4.

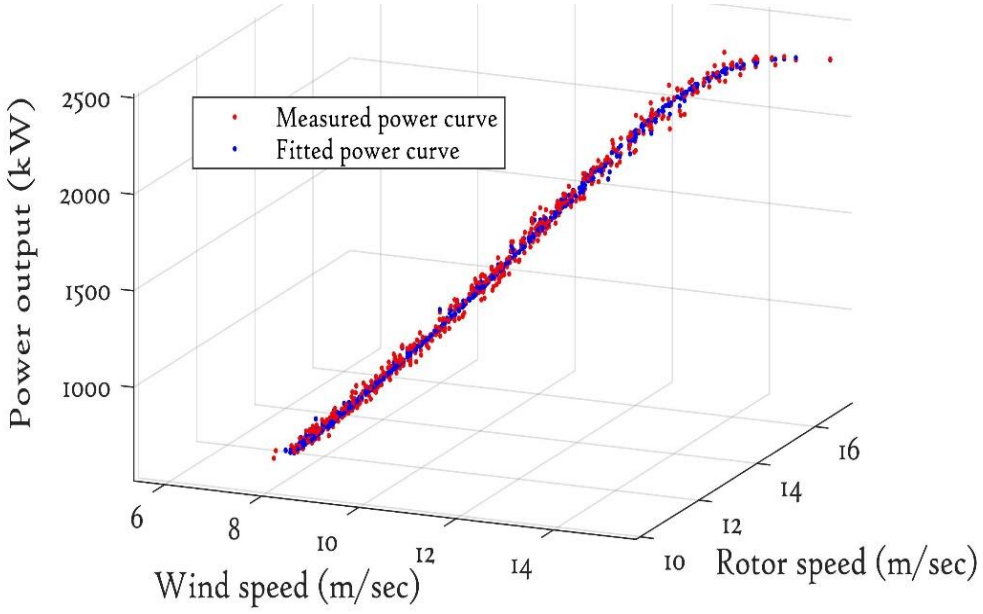


Figure 8.4: Estimated GP power curve with the inclusion of rotor speed

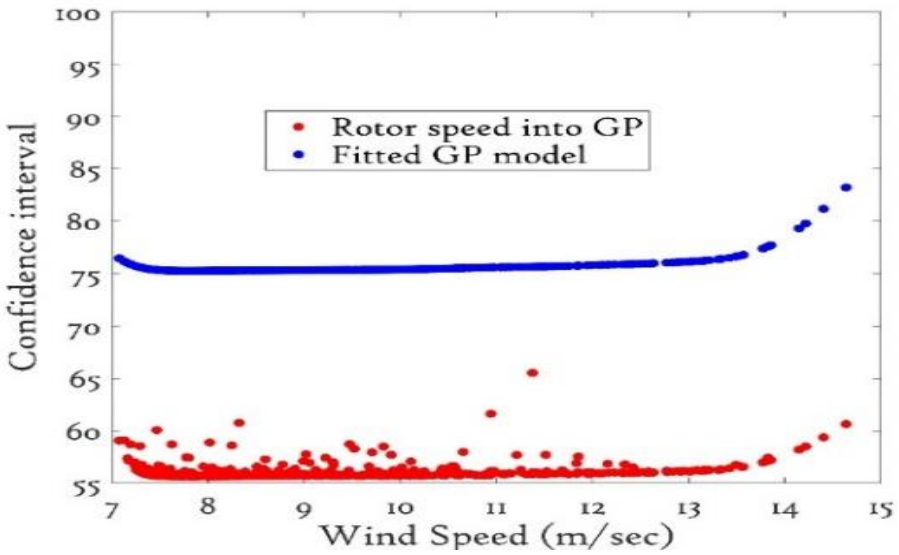


Figure 8.5: GP models Uncertainty analysis when rotor speed incorporated

The estimated CIs values are plotted as a function of wind speed for GP model uncertainty analysis purposes as in Figure 8.5. The GP power curve model uncertainty reduces significantly across the wind speed range when the rotor speed is incorporated into the model as shown in Figure 8.5. This improvement in GP model uncertainty is significant for constructing robust GP fault detection algorithms and will be discussed in great details in upcoming sections.

8.3 Comparative studies of the impact of wind turbine performance parameters on Gaussian Process models accuracy

In this section, a performance comparison of GP models is presented in which uncertainty analysis and performance error metrics are used to judge the impact of additional performance parameters on GP model accuracy. Additionally, the impact of the inclusion of both additional performance parameters on the GP models is investigated.

Figure 8.6 depicts the relationship of estimated CIs against wind speed for separate and combined inclusion of the additional parameters. The results show that the GP models with the inclusion of rotor speed make a significant improvement on model uncertainty as compared to the GP model with and

without the inclusion of blade pitch angle, see Figure 8.6. Furthermore, when GP with the inclusion of rotor speed model uncertainty is compared with the GP with inclusion blade pitch angle and rotor speed together, then the improvement in uncertainty is almost the same in both models. This confirms that incorporating rotor speed alone to GP models makes a compelling improvement which can be useful for constructing robust GP fault detection algorithms.

Statistical error metrics were used here to confirm the impact of these performance variables on GP model accuracy. The statistical error metrics used to evaluate the performance of nonparametric models are MAE, RMSE, and coefficient of determination (R^2); described briefly in chapter 3. The calculated values of these error metrics are tabulated in Table 8.1. The calculated RMSE and MAE values reflect the amplitude errors and their small values indicate a strong relationship between measured and predicted values. In the GP model with rotor speed, RMSE and MAE values are smallest which suggests that a GP model incorporating rotor speed showed significant accuracy improvement. Furthermore, ithe nclusion of rotor speed gives R^2=0.9969, which makes GP model relatively very close to 1 as compare to other performance parameters. Hence, highlighting the significant impact of rotor speed on GP model. It is worth noting that the GP model with inclusion of rotor speed and pitch angle have the highest accuracy but its value close to GP with rotor speed model, see Table 8.1. This trend well agrees on the results in Figure 8.6.

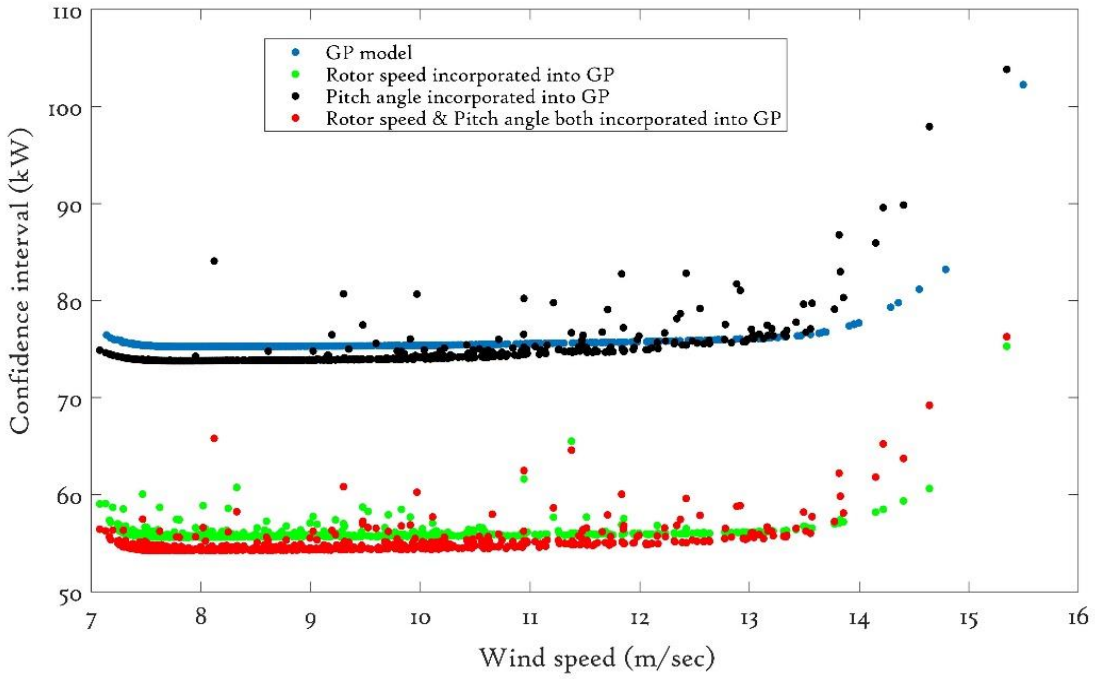


Figure 8.6: Comparative analysis of the impact of performance parameters on GP model uncertainty.

GP Models	RMSE	R^2	MAE
without inclusion	38.04	0.9942	28.96
with the inclusion of rotor speed.	27.89	0.9969	21.17
with the inclusion of pitch angle	37.04	0.9945	28.22
with the inclusion of pitch angle, and rotor speed	27.03	0.9971	20.61

Table 8.1: Statistical measures for GP models when performance parameter included

8.4 Gaussian Process fault detection algorithm with the inclusion of rotor speed

Results from the previous section suggest that rotor speed makes a significant improvement to GP model accuracy and reduces uncertainty. This improvement can be useful in developing a robust GP based fault detection algorithm. To validate this, a fault detection GP algorithm is constructed and compared with that of Chapter 7. The performance comparison between these two algorithms is

undertaken concerning their capability to detect in advance (and by how much) signs of failure, and also their false positive rate. Besides impacting the power producing ability of a turbine, yaw error also affects the reliability of critical subsystems in wind turbines as was discussed in chapter 7. This motivates, as before, consideration of yaw misalignment as a case study to validate the impact of rotor speed on the GP fault detection algorithm.

Dataset	Start timestamp	End timestamp	Description
1	11/1/2008 14:30 PM	30/03/2008 15:20 PM	Total data filtered set: 3274 observations
2	14/4/2009 11:20 AM	16/4/2009 9:50 AM	Total data set: 201 observations

Table 8.2: Description of the SCADA datasets

SCADA data starting from 11/1/2008 14:30 PM until 30/03/2008 15:20 PM is used to train the GP power curve model as shown in Table 8.2. It should be noted that the SCADA data used are filtered, and air density corrected as per the methodology described in chapter 5. The reference GP power curve was constructed using GP theory with rotor speed incorporated in a squared exponential covariance function of GP model as per the procedure described in section 8.3; the result is shown in Figure 8.7. The obtained GP power curve is compared with a measured power curve and found that GP effectively represents the expected variance (i.e. the spread of the data in the power curve). It should be noted again that Figure 8.7 is the air density corrected power curve.

The solid blue lines are the estimated CIs associated with the GP power curve model; it is clear that rotor speed inclusion results in CIs with significantly narrowed spread (as compared to the chapter 7 GP reference power curve). This results in an improvement in GP model accuracy by rejecting the data points effectively that lie outside of the CIs. The rejected data points can be considered anomalous, signifying a potential malfunction of the wind turbine.

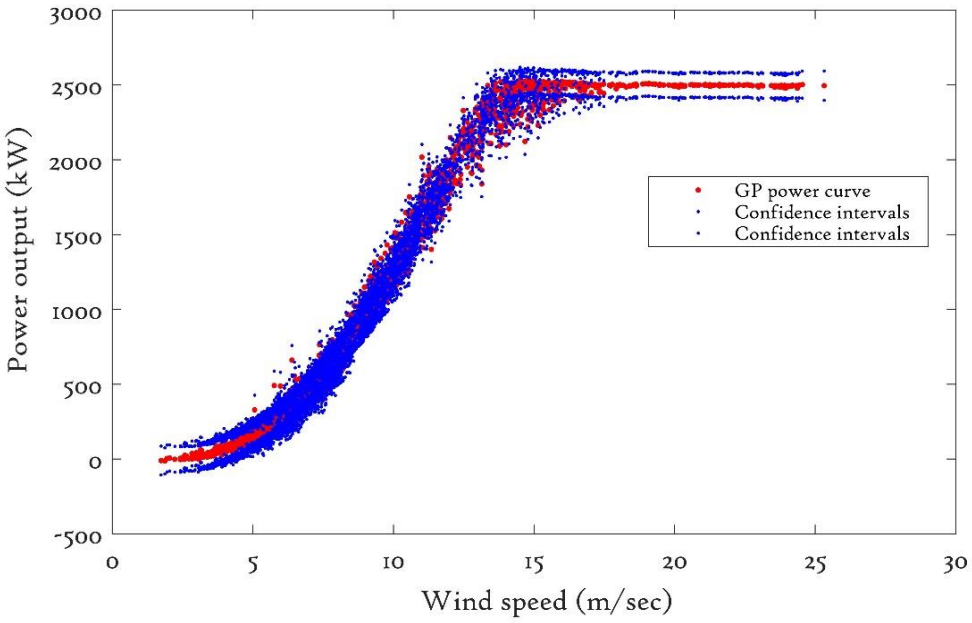


Figure 8.7: GP power curve with CIs when the rotor speed is incorporated

SCADA dataset 2 of Table 8.2 represent an unhealthy dataset where significant yaw error occurred when the turbine failed to yaw for an extended period despite changes in wind direction, as illustrated in Figure 7.6 (of chapter 7).

The incoming unhealthy data points of datasets 2 are judged on a point by point basis against the reference GP power curve (Figure 8.7), and probabilistic assessment was undertaken (as described in chapter 7). The GP model is fixed at the probability threshold of 0.006 to prevent GP algorithm from generating false alarms and therefore saving the additional operational cost caused by a false alarm. The threshold of 0.006 gives an accurate result without false positives, and thus it is used here. Using this, an alarm would have been raised at 21:50 on 14/04/2009, just 50 minutes after the start of the yaw fault at 21:00 on 14/04/2009, without any false positives; this is illustrated in Figure 8.8.

Note that chapter 7 GP fault detection algorithm detects the failure at 22:40 on 14/4/2009, ~1.5 hours after the start of the yaw fault at 21:00 on 14/04/2009. The GP model with the inclusion of rotor speed when compared with this algorithm was able to detect the first sign of anomaly 40 minutes earlier than the model without rotor speed as tabulated in Table 8.3. This confirms that the inclusion of

rotor speed improves the GP fault detection algorithm; in particular, its capability to detect failure quickly without any false alarm.

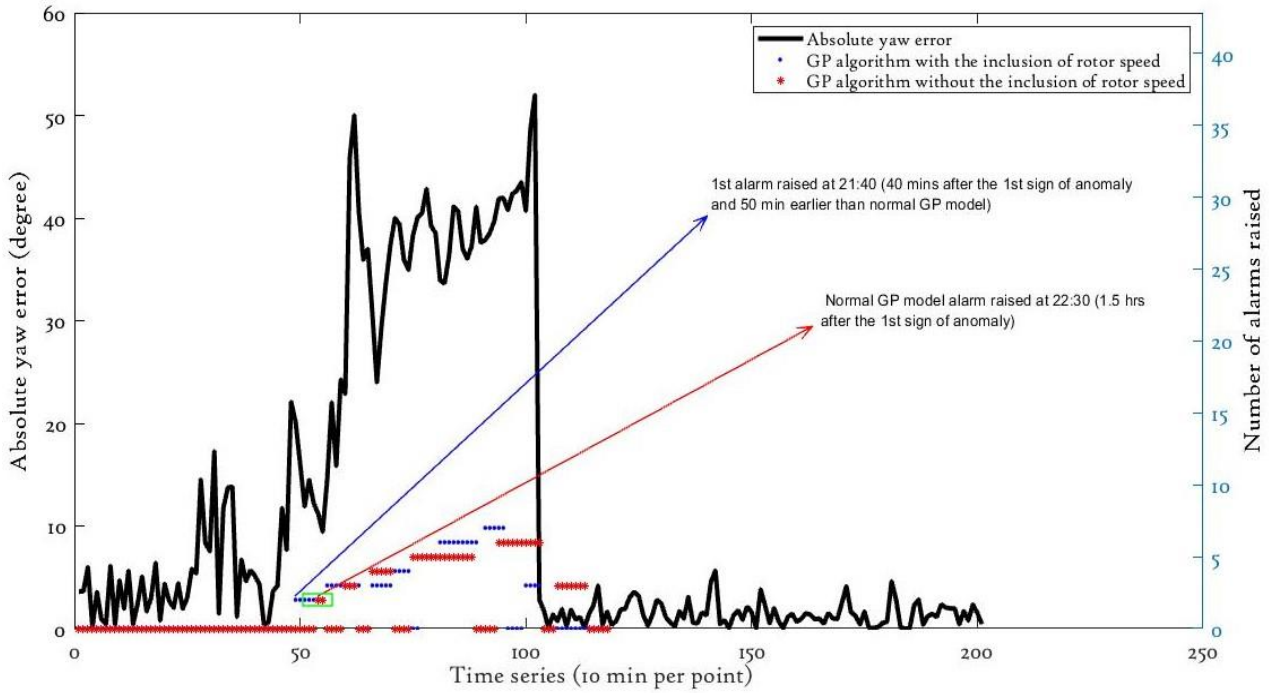


Figure 8.8: Impact of rotor speed on GP absolute yaw error detection algorithms

GP Models	Alarm detected	Time is taken to identify the fault
Probabilistic assessment using GP	22:40 on 14/4/2009	~ 1.5 hours
Probabilistic assessment using GP with the inclusion of rotor speed	21:50 on 14/4/2009	50 minutes

Table 8.3: Alarm record and detection by GP models when rotor speed incorporated

It is worth to note that in chapter 7, GP fault detection algorithm used the Fisher Test in which 3 p-values with a threshold of 0.008 applied to filter the individual p-values. However, the inclusion of rotor speed narrows down the width of CIs (Figure 8.7), and due to this at a threshold value of 0.006, the obtained p-value is high. So, there no need to use 'Fisher Test' since its p-value is high unlike the p-values obtained in chapter 7.

8.5 Chapter conclusions

The power curve accuracy is affected by various internal and external factors and taking these factors into account improves the model accuracy and its capabilities to detect the failures. In this chapter, the impact of additional wind turbine performance parameters (pitch angle and rotor speed) on GP power curve model accuracy and uncertainty was explored. The GP power curve was developed by incorporating these turbine performance parameters into the model. The inclusion of rotor speed improves the GP model accuracy and uncertainty which is further verified by performance error metrics (Table 8.1). Though the inclusion of blade pitch angle also improves the GP model accuracy somewhat, compared to rotor speed, it is not significant as presented in Figure 8.6 and Table 8.1.

A fault detection algorithm based on the GP approach extended to include rotor speed was tested against the GP fault detection algorithm of chapter 7 in order to assess the impact of rotor speed on algorithm fault detection capabilities. Yaw misalignment which causes significant power loss was used here as a case study against which algorithm was tested and evaluated. Comparative Studies of these two methods concludes that inclusion of rotor speed improves the GP fault detection capabilities and able to detect the failures 40 minutes earlier than chapter 7 fault detection algorithm, see Figure 8.8 and Table 8.3.

Chapter 9

Conclusions, discussion and future works

The offshore wind industry is forecast to experience significant growth over the coming years and will continue to play a significant role in providing electricity to the UK households and industries. However, offshore turbines attract higher O&M costs than those onshore due to the logistics, transportation and high maintenance costs. Unscheduled maintenance due to unexpected failure can cause high O&M costs, and early detection of these failures that can help prevent catastrophic failures, improve turbine availability, reduces O&M cost, and improve the power performance of wind turbines.

A Gaussian Process is a data-driven approach that has demonstrated effective application in many research areas. However, its application to wind turbine condition monitoring has to date been somewhat limited. This chapter summarises the main research contributions of this thesis in context to Gaussian Process model application to wind turbine condition monitoring and is followed by an outline of potential future work.

9.1 Summary of thesis contributions and the overall conclusion

The main contribution of this thesis has been to demonstrate the potential for SCADA data based Gaussian Process models for wind turbine condition monitoring. Of the total of 9 chapters, chapters 3 to 8 present specific scientific contributions to this field and these are summarised as follows.

- **Key points from Chapter 3**

The covariance (or kernel) function is the key to the GP model and is used to signify the similarity between two given points and measure the degree to which both points are related. Different forms of covariance function are available, and

hence the selection and composition of covariance functions is a non-trivial task in GP modelling. This comparison is essential for robust GP fault detection algorithm development and needs to be done at the first stage so that based on this analysis; a suitable covariance function can be selected for future research. So, in this chapter in-depth analysis of popularly used stationary covariance functions are taken into GP power curve modelling and a comparative analysis undertaken. The commonly used squared exponential covariance function is taken as the benchmark, against which to assess the other covariance functions. The constructed GP models are validated using SCADA datasets from healthy operational wind turbines.

The results show that fitted power curve GP model based on rational quadratic accuracy is almost same as the commonly used squared exponential covariance functions although prediction speed and time is taken by model training is extended, whereas GP models based on the squared exponential covariance function performance are superior. In short, both covariance functions worked pretty well in identifying a well fitted smooth function. The rational quadratic covariance function can be used instead of squared exponential covariance functions for GP modelling if the data sets are not large and there is no limitation on training time and prediction speed. In contrast, the Matern 5/2 covariance function should be used with caution, and GP models based on the Matern 3/2 function perform poorly and are not recommended for wind turbine power curve modelling.

Based on the result from Chapter 3, the squared exponential is found to be the most suitable covariance function for power curve modelling and was used in all subsequent chapters.

- **Key points from Chapter 4**

In this chapter, three advanced nonparametric approaches, namely Gaussian Process (GP), Random Forest (RF), and Support Vector Machine (SVM) are

assessed for wind turbine power curve modelling. GP and SVM are kernel based methods while RF is based on a regression tree and is now starting finding application in wind turbine condition. The results obtained from these methods have been compared using suitable performance error metrics to identify the best method for modelling of the power curve and to list out the strengths and weaknesses of models.

The computational results and the analysis (chapter 4) have demonstrated that the power curves based on GP models have the highest fitting accuracy, and can reflect the dynamic properties of a power curve, whereas the SVM approach gives a sufficiently accurate result but within a restricted wind speed range. Power curves based on GP and SVM provide a smooth and continuous fitting, whereas the power curve based on the RF is neither smooth nor continuous. The confidence intervals come with estimated mean values with GP models which makes uncertainty analysis straightforward while in case of RF and SVM, the additional mathematical computation required for confidence interval calculation increases the mathematical computational challenges and extra cost as compared to GP models.

- **Key points from Chapter 5**

The wind turbine power curve is mostly used to assess wind turbine performance, but it is not a perfect indicator because various failures and downtime events may go undetected if only this is used. Therefore, it is desirable to explore other curves that are based on critical parameters that affect the operation and power production of the wind turbine. Wind turbines operation is affected by external factors (e.g., wind turbulence, wind shear and icing) and internal factors (e.g., rotor power, rotor speed and blade pitch angle). The external factors cannot be controlled, but the events related to internal factors can be analysed and potentially controlled to optimise the performance of wind turbines. The internal operations of turbines depends on critical variables, in particular, rotor power,

rotor speed and blade pitch angle, and continuous monitoring of these parameters improves the overall effectiveness of any model used to maximise the power production of turbines.

In chapter 5, two GP operational curves, namely the rotor speed curve and blade pitch angle curve are developed alongside the power curve for delivering robust SCADA based condition monitoring. The constructed GP operational curves are compared with the conventional approach based on binned operational curves together with individual bin probability distributions to identify operational anomalies. Finally, a comparative analysis of these techniques is presented and explained how these GP operational curves could be used for effective condition/performance monitoring.

- **Key points from Chapter 6**

Power output is well known to be influenced by air density, and this is reflected in the IEC Standard air density correction procedure. Chapter 6 presented and assessed four different possible air density compensation approaches and the primary objective of this research is to explore whether the IEC (traditional approach) to air density correction is the most effective when estimating power curves using GP models.

The result suggests that adding density to the GP without any pre-correction makes the GP model more effective and reduces uncertainty significantly as compared to other approaches, as validated here by distribution functions, confidence intervals, and statistical measures analysis. In short, not applying the Standard IEC density correction, but instead including air density in the GP model gives significantly more accurate power curves.

- **Key points from Chapter 7**

Chapter 7 demonstrates the application of a GP algorithm to wind turbine anomaly detection. The specific fault investigated was a yaw control failure. In order to judge GP model effectiveness, two other methods based on binning have

been tested and compared with the GP based method. It was found that the GP model was able to detect the anomaly effectively with the alarm raised only 1.5 hours after the fault occurred. Not only is the GP method developed ability to detect the yaw misalignment quickly, but it produces no false positives, in contrast to the bin based models, hence confirming that the GP approach provides both fast and robust fault identification.

- **Key points from Chapter 8**

The electrical power not only correlates with wind speed but also reflects various turbine parameters such as rotor speed and blade pitch angle. In this chapter, these parameters are incorporated into the GP model in order to analyse its impact on accuracy and uncertainty. The result demonstrated that inclusion of rotor speed makes a significant improvement in GP model accuracy and uncertainty. This improvement used to improve the GP fault detection algorithm of Chapter 7. The comparative studies of the two GP algorithm showed that the inclusion of rotor speed in the GP model increased its capabilities to detect the failure quickly, in fact, the extended model takes only 50 minutes to detect the first sign of failure.

9.2 Future works

This thesis provides an exciting opportunity to further advance the use of Gaussian Process models in context to wind turbine condition monitoring based on SCADA data. However, this thesis work can also be extended in many directions as outlined below.

- **Extending this research**

In this thesis, a GP fault detection algorithm was tested against a fault resulting in yaw misalignment. The constructed algorithm should also be tested with a range of failures that results in turbine underperformance.

In most of this thesis work, the GP power curve was used to assess the underperformance of a wind turbine. However, it has been suggested demonstrated in chapter 5 that many downtime events and failures remain

undetected by the power curve. This should lead to an examination of other key indicators that could act as an add-on to the power curve for robust wind turbine condition monitoring. GP based reference blade pitch curve and rotor curves could be used along with power curve to assess the power performance of wind turbine effectively. In future work, the developed GP operational curves of chapter 5 could be applied to detect downtime events which remains undetected by the power curve. Also, two generators for redundancy and how can GP model distinguish between the two different power curves are also included in future works.

- **Improving GP model accuracy and reducing uncertainty**

Chapter 8 demonstrated that additional wind turbine performance parameters improve the GP fault detection algorithm accuracy significantly. Likewise, other parameters should be investigated such as turbulence intensity, wind shear and so on, that are known to affect the power production of wind turbines. A comparative analysis of such parameters may improve GP model accuracy and reduce uncertainty and thus merit attention.

- **Performance comparison with existing statistical techniques**

This thesis gave a basic foundation of the GP model for SCADA data based wind turbine condition monitoring. The comparative studies of fault detection algorithms (binning) demonstrated that the capabilities of GP models to detect the failures at its early stage with any false positives. However, the constructed GP fault detection algorithm needs to compare with other available statistical models (e.g. SVM and RF), and this kept for future work.